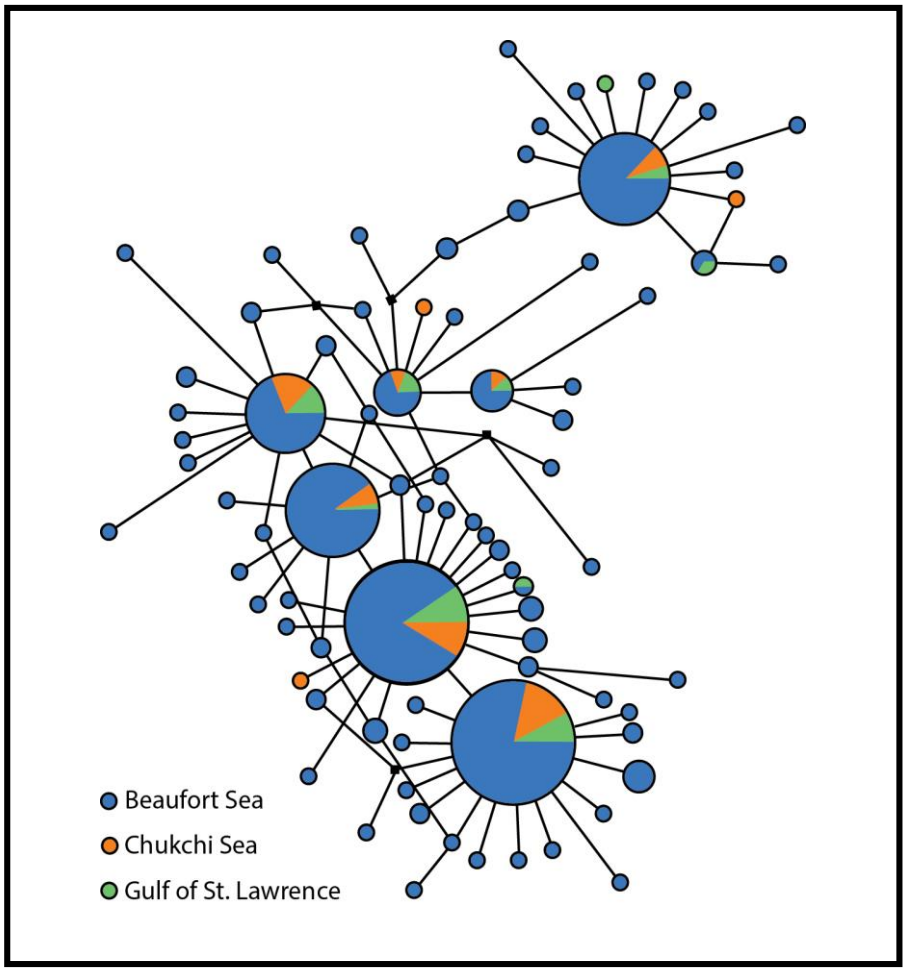
OCS Study
BOEM 2017-066

Genomics of Arctic Cod



US Department of the Interior
Bureau of Ocean Energy Management
Alaska OCS Region

BOEM
Bureau of Ocean Energy Management

OCS Study
BOEM 2017-066

Genomics of Arctic Cod

30 September 2017

Authors

Robert E. Wilson, George K. Sage, Sarah A. Sonsthagen, Meg C. Gravley, Damian M. Menning and Sandra L. Talbot

Prepared under BOEM Award

M14PG00008

By

U. S. Geological Survey, Alaska Science Center

4210 University Drive

Anchorage, AK 99508 USA

US Department of the Interior
Bureau of Ocean Energy Management
Alaska OCS Region

DISCLAIMER

This study was funded, in part, by the US Department of the Interior, Bureau of Ocean Energy Management (BOEM), Environmental Studies Program, Washington, DC, through Agreement Number M14PG00008 with the U.S. Geological Survey (USGS). The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the Bureau of Oceans Energy and Management.

REPORT AVAILABILITY

To download a PDF file of this report, go to the U.S. Department of the Interior, Bureau of Ocean Energy Management Data and Information Systems webpage (https://marinecadastre.gov/espis/#) and search for 2017-066.

CITATION

Wilson, R.E, G.K. Sage, S.A. Sonsthagen, M.C. Gravley, D.M. Menning, and S.L. Talbot. 2017. Genomics of Arctic Cod. Anchorage, AK: US Dept. of the Interior, Bureau of Ocean Energy Management, Alaska OCS Region. OCS Study BOEM 2017-066. 92pp.

ACKNOWLEDGEMENTS

We thank the numerous staff at U.S. Geological Survey for providing samples and for laboratory and graphical assistance. The BOEM Alaska OCS Region contributed to this document. We thank Brenda Norcross (University of Alaska Fairbanks), Vanessa von Biela (USGS), Andrew Majewski (Fisheries and Ocean Canada), Tony Gharrett (University of Alaska Fairbanks), Noel Sme (University of Alaska Fairbanks), Lorena Edenfield (University of Alaska Fairbanks), and John Nelson (University of Victoria, Canada), University of Washington Burke Museum, Institute of Marine Science at the University of Alaska, Bureau of Ocean Energy Management, and Fisheries and Ocean Canada for providing samples. Gabriel DeGange (USGS) provided laboratory support. We also thank Han Ming Gan (Deakin University) for mitogenome assembly support. This report has been reviewed and approved for publication consistent with USGS Fundamental Science Practices. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Table of Contents

List of Figures

List of Tables

List of Supplementary Tables

List of Abbrevations Acronyms and Symbols

AHR	Aryl hydrocarbon receptor
BLAST Basic	Local Alignment Search Tool
BOEM	Bureau of Ocean and Energy Management
bp	base pairs
CO1	cytochrome oxidase 1
cytb	cytochrome b
ddRAD-Seq	Double digest Restriction Site Associated DNA Sequencing
HWE	Hardy-Weinberg Equilibrium
LD	linkage disequilibrium
MHC	Major Histocompatibility Complex
mtDNA	mitochondrial DNA
NGS	Next Generation Sequencing
ND	NADH dehydrogenase
OCS	Outer Continental Shelf
OXPHOS	oxidative phosphorylation
LR-PCR	Long-Range polymerase chain reaction
RAD-Seq	Restriction site Associated DNA Sequencing
RNA-Seq	RNA Sequencing
USGS	U.S. Geological Survey

Executive Summary

The Arctic cod (*Boreogadus saida*) is an abundant marine fish that plays a vital role in the marine food web. To better understand the population genetic structure and the role of natural selection acting on the maternally-inherited mitochondrial genome (mitogenome), a molecule often associated with adaptations to temperature, we analyzed genetic data collected from 11 biparentally-inherited nuclear microsatellite DNA loci and nucleotide sequence data from from the mitochondrial DNA (mtDNA) cytochrome b (cytb) gene and, for a subset of individuals, the entire mitogenome. In addition, due to potential of species misidentification with morphologically similar Polar cod (*Arctogadus glacialis*), we used ddRAD-Seq data to determine the level of divergence between species and identify species-specific markers.

Based on the findings presented here, Arctic cod across the Pacific Arctic (Bering, Chukchi, and Beaufort Seas) comprise a single panmictic population with high genetic diversity compared to other gadids. High genetic diversity was indicated across all 13 protein-coding genes in the mitogenome. In addition, we found moderate levels of genetic diversity in the nuclear microsatellite loci, with highest diversity found in the Chukchi Sea. Our analyses of markers from both marker classes (nuclear microsatellite fragment data and mtDNA cytb sequence data) failed to uncover a signal of microgeographic genetic structure within Arctic cod across the three regions, within the Alaskan Beaufort Sea, or between near-shore or off-shore habitats. Further, data from a subset of mitogenomes revealed no genetic differentiation between Bering, Chukchi, and Beaufort seas populations for Arctic cod, Saffron cod (*Eleginus gracilis*), or Walleye pollock (*Gadus chalcogrammus*). However, we uncovered significant differences in the distribution of microsatellite alleles between the southern Chukchi and central and eastern Beaufort Sea samples of Arctic cod. Finally, using ddRAD-Seq data, we identified species-specific markers and in conjunction with mitogenome data, identified an Arctic cod x Polar cod hybrid in western Canadian Beaufort Sea.

Overall, the lack of genetic structure among Arctic cod within the Bering, Chukchi and Beaufort seas of Alaska is concordant with the absence of geographic barriers to dispersal and typical among marine fishes. Arctic cod may exhibit a genetic pattern of isolation-by-distance, whereby populations in closer geographic proximity are more genetically similar than more distant populations. As this signal is only found between our two fartherest localities, data from populations elsewhere in the species' global range are needed to determine if this is a general characteristic. Further, tests for selection suggested a limited role for natural selection acting on the mitochondrial genome of Arctic cod, but do not exclude the possibility of selection on genes involved in nuclear-mitogenome interactions. Unlike previous genetic assessment of Arctic cod sampled from the Chukchi Sea, the high levels of genetic diversity found in Arctic cod assayed in this study, across regions, suggests that the species in the Beaufort and Chukchi seas does not suffer from low levels of genetic variation, at least at neutral genetic markers. The large census size of Arctic cod may allow this species to retain high levels of genetic diversity. In addition, we discovered the presence of hybridization between Arctic and Polar cod (although low in frequency). Hybridization is expected to occur when environmental changes modify species distributions that result in contact between species that were previously separated. In such cases, hybridization may be an evolutionary mechanism that promotes an increase in genetic diversity that may provide species occupying changing environments with locally-adapted genotypes and, therefore, phenotypes. Natural selection can only act on the standing genetic variation present within a population. Therefore, given its higher levels of genetic diversity in combination with a large population size, Arctic cod may be resilient to current and future environmental change, as high genetic diversity is expected to increase opportunities for positive selection to act on genetic variants beneficial in different environments, regardless of the source of that genetic variation.

1 Introduction

Arctic cod are considered a keystone forage fish species in the U.S. Arctic marine ecosystem and are estimated to funnel 75% of lower trophic energy to upper trophic marine and near-shore predators including birds, seals, beluga whales (*Delphinapterus leucas*), and eventually to polar bears (*Ursus maritimus*) and humans (Bradstreet et al. 1986, Welch et al. 1992, Crawford and Jorgenson 1996). Information about the species is needed to inform Essential Fish Habitat (EFH) related National Environmental Policy Act (NEPA) analyses. However, the absence of basic information on distribution and population patterns of Arctic cod (see Mueter et al. 2016), precludes complete understanding of the status of the Arctic cod in northern waters the speices likey response to perturbations.

Genetic diversity is considered to be one of three levels of biodiversity warranting conservation by the World Conservation Unit (McNeely et al. 1990). Understanding the distribution of genetic variation within a species – that is, defining populations and searching for population boundaries – and determining the level of genetic diversity within a population is important for resource managers since it allows them to predict the impact of environmental challenges within a particular locale. Environmental challenges might impact each population differently, and it is well known that smaller populations and population characterized by low levels of genetic diversity are more vulnerable to environmental impacts than larger populations characterized by high genetic diversity (Reed and Frankham 2003).

Thus, a single, large panmictic population of Arctic cod characterized by moderate to high levels of genetic diversity might be expected to be more resilient to environmental perturbations than would Arctic cod that belong to smaller, fragmented populations each characterized by low genetic divesity. Information that can be provided by genetics/genomics analyses of Arctic cod include 1) clarifying whether the Arctic cod in the Outer Continental Shelf (OCS) region comprise a single, panmictic population; and 2) determining whether Arctic cod possess genetic characteristics that enable them to adapt to a changing marine environment.

An early analysis by the USGS and BOEM (Talbot et al. 2014) found evidence for population differentation between Arctic cod occupying the central Beaufort Sea and the Chukchi Sea, based on nucleotide sequence data from the mtDNA cytb gene, a gene that carries signals of more historical processes than other markers, such as microsatellite markers, that are used to reflect more recent processes. The presence of these two divergent mitochondrial lineages in Talbot et al. (2014) indicates the potential of admixture through dispersal beween previous segregated groups or a stable population size during the last glacial period. Similarly, Pálsson et al. (2009) found little evidence of geographic subdivision in Arctic cod, based on sequence data from the mitochondrial DNA (mtDNA) cytochrome b (cytb) gene from cod sampled across 19 locations, including Greenland, Iceland, Svalbard, and the Bering Strait (Chukchi Sea). Similarly, Wildes et al. (2016) found no population-level structuring within the Chukchi Sea, based on sequence data from the mtDNA cytochrome oxidase 1 gene and fragment data from 15 microsatellite loci. Thus, Arctic cod of the Beaufort Sea are assumed to form a single panmictic population and not be structured into subgroups or subpopulations.

The Arctic cod is well suited for life in sub-zero Celsius water temperatures, in part due to the evolution of one of the four structurally diverse antifreeze proteins (Chen et al. 1997, Pörtner and Playle 1998). However, some researchers hypothesize that Arctic cod may have difficulties adapting to warmer temperatures, and be subject to interspecies competition (e.g. with other gadids and haddock (*Melanogrammus aeglefinus*; Renaud et al. 2012), given the predicted influx of more temperate-adapted species (Nahrgang et al. 2014, Drost et al. 2016, Laurel et al. 2016). Recent climate-change modeling suggests that as the Arctic sea ice retreats, Arctic cod may be extirpated in most of its range within the next 30 years (Cheung et al. 2008). A greater understanding of genomics of Arctic cod may also provide insight

in the ability of the species to adapt (adaptive capability) as sea ice retreats. A combination of classical genetics, genomics, and transcriptomics technologies can provide insight into whether certain Arctic cod lifestages are truly ice-obligate or whether they are simply ice-associated and can potentially adapt to retreating ice conditions. Arctic cod occupy a wide breath of habitats as they can be found throughout the water column as well as in near-shore shallow water to hundreds of kilometers off-shore (Thorsteinson and Love 2016). However, little is known about the seasonal movements within Alaskan waters. Assessments of levels of genetic partitioning coupled with data from the transcriptome, particularly among near-shore and off-shore populations, could provide insight into whether differential adaptability is due to differential gene expression rather than different genes or if distinct sub-populations are utilizing different habitat types. The transcriptome presented in this study can be used in to digital gene expression panels that can be leveraged under controlled experimental conditions to uncover differential expression of the single 'antifreeze' gene that may confer adaptability to loss of ice habitat. Further, genomic coupled with transcriptomic technologies using samples in controlled environments can be used to identify additional genes associated with possible ice-obligation.

The main footprint of this study ranges from U.S. Beaufort Sea across the Mackenzie River Canyon into the Western Canadian Beaufort, which coincides with the footprint of Norcross et al. (2016). That study collected genetic samples for analysis by the USGS Alaska Science Center, which has led the Arctic cod genomics/transcriptomics research. Other samples include the following planning areas: Chukchi Sea, Hope-Basin, Norton Basin, St. Matthew-Hall, Navarin Basin, Aleutian Arc, St. George Basin, and Kodiak.

This report is divided into three sections: (1) introductory material, (2) population structure and differentiation across the study area, comparative phylogeography of Arctic cod and other gadids, and specis identification and potential speices hybridization, and (3) presentation of a transcriptome resource for Arctic cod that may serve as a baseline for future adaptive genetic research.

1.1 Objectives

This study provides genetic and genomic products that tested the following hypotheses (Section 2) and a transcriptomic resource (Section 3) intended to facilitate future research:

- Test hypotheses related to possible population genetic structure that would inform BOEM as to whether Arctic cod are a single panmictic population or comprise several geographic sub-populations.
- Test whether there is micro-geographic genetic structure within the Beaufort Sea which may relate to near-shore/off-shore differences.
- Test hypotheses of natural selection acting on Arctic cod mitogenome compared to other gadids found in Alaskan waters.
- Obtain species level loci to properly identify Arctic cod and Polar cod and investigate level of hybridization.
- Provide an archive of genetic and transcriptomic data that can be used for future research on the adaptability of Arctic cod to a changing marine ecosystem.

1.2 Notes on Terminology

It is important to clarify certain terms that are used throughout the report, including the following: *genetics* (including 'classical genetics'), *genomics, mitogenome, transcriptomics*, *population*, *positive, negative* and *balancing selection*, and *genetic diversity*.

Genomics is an overarching, interdisciplinary scientific field that attempts to characterize and quantify the entire genome, the entire set of DNA within a single cell of an organism. Vertebrates have two genomes – the nuclear genome, which is inherited from both parents, and the mitochondrial genome (the *mitogenome*), which is inherited from only the mother. Genomics research focuses on the description of the structure, function and evolution of the genome, in addition to the mapping of genomes via sequencing. *Reduced representational genomics* focuses on the use of a subset of genes to represent the genome, and can utilize both neutral and functional changes in the genome to make inference about differences between groups.

Neutral genetic variation, or variation that does not affect fitness and therefore the evolutionary fate of genetic variation, is the purview of 'classical genetics.' *Classical genetics* uses reduced representational genomics panels that assay widely used neutral marker classes, such as variation in nucleotide sequences of the maternally-inherited vertebrate mitogenome or fragment analysis of biparentally-inherited multiallelic microsatellite loci (genotyping), in order to determine the genealogical relationships among individuals, populations, metapopulations, species, and higher taxonomic levels. For the past several decades, classical genetics analyses of vertebrate species have largely leveraged nucleotide sequence data from genes in the mitogenome, and fragment data from biparentally-inherited nuclear microsatellite loci. More recently, reduced representational genomic analyses have concentrated on the assessment of variation in *single-nucleotide polymorphisms*, or SNPs, using neutral SNPs to delineate population boundaries, or functional SNPs — SNPs that reflect mutational differences that may impact the fitness of an organism — to delineate functional differences between groups (populations, species, etc.). While genomics involves the assessment of both neutral and functional genes, *transcriptomics* is the study of the transcriptome, the complete set of RNA transcripts produced by the genome, under particular circumstances or in a specific cell.

Within this report, we define a *population* using a biological definition: a population is a group of organisms within one species that interbreed and live in the same place at the same time. Population genetics/genomics analyses are intended to test whether a group of individuals within an aggregation actually comprise a biological population, whether they are a subset of a larger population, or whether they are subdivided into subgroups or subpopulations with varying levels of evolutionary dispersal, or gene flow. To avoid conflating divergence by descent (that is, divergent evolution) with convergence by function (that is, convergent evolution), population structure is assayed using neutral genetic markers, as appropriate for studies seeking to delineate population or taxonomic boundaries. Typically, the null hypothesis in such tests is that the particular aggregation under study represents a single, panmictic population and the resulting test either supports or rejects that null hypothesis. Further exploratory investigations serve to either confirm the intitial outcome, or seek for trends that are not significant, but may be important to understanding the evolutionary biology of the species.

Genetic diversity can simply be thought of as the amount of variation in the genetic makeup within an individual, population, or species. There are both evolutionary and ecological consequences associated with levels of genetic diversity (Hughes et al. 2008). As environmental factors change, genetic diversity is needed for populations to be able to adapt or response to these new conditions (Reed and Frankham 2003). In other words, genetic diversity is directly associated with a species "adaptive capacity" as it provides the building blocks for natural selection to act upon, and if a population has low levels of genetic diversity, its ability to respond to changes in environmental conditions will likely also be limited. In addition, a population's fitness/survival is often tightly linked to the level of heterozygosity, a common measure of the degree of genetic variation (Vandewoestijne et al. 2008, Markert et al. 2010). For example, a reduction in genetic diversity may make a population more vulnerability to disease (O'Brien and Evermann 1988). Genetic variation can be produced or maintained through a variety of processes, including mutation (a change in an organism's DNA), random mating, random fertilization, recombination or crossing over (exchange of genetic material between homologous chromosomes),

natural selection, and gene flow, including hybridization, via the incorporation of new genetic combinations into the population. Natural selection, which operates largely on functional and regulatory genes, can be positive, negative or balancing. *Positive selection*, which is also called Darwinian selection, is the process by which new, advantageous genetic variants sweep a population. *Negative selection*, also called purifying selection, is the selective removal deleterious alleles, resulting in stabilizing selection maintained via the elimination of alleles that reduce fitness. *Balancing selection* refers to a suite of selective processes by which multiple versions of a gene are actively maintained in a gene pool at frequencies higher than expected due to genetic drift alone.

1.3 Notes on Taxonomy

Boreogadus saida is often referred to as both Arctic cod and Polar cod in the scientific literature. The use of multiple common names has led to confusion, since the species *Arctogadus glacialis* is also referred to both Arctic cod and Polar cod. Recently, the American Fisheries Society and American Society of Ichthyologists and Herpetologists have recommended using Polar cod for *Boreogadus saida* to follow European taxonomic nomenclature (Page et al. 2013). However, in this report, we use the common name "Arctic cod" to refer to *Boreogadus saida* and "Polar cod" to refer to *Arctogadus glacialis*; this allows us to remain consistent with the majority of Alaskan literature. In addition, we use the common names "Saffron cod" and "Walleye pollock" throughout this report to refer to *Eleginus gracilis* and *Gadus chalcogrammus*, respectively.

1.4 Study Chronology

We commenced analyses of Arctic cod with a pilot study (BOEM, 2014-050; Talbot et al. 2014) in 2011, followed by classical genetics studies of nucleotide data from the mtDNA cytb gene using additional samples collected in 2011-2014. The goal of the pilot study was to develop and test molecular markers and determine levels of genetic structure between Arctic cod sampled in the Chukchi and western Beaufort seas to those from a geographically distinct population in Nova Scotia, Canada. Beginning in 2014, John Nelson of University of Victoria, Victoria, British Columbia, and other researchers began laboratory analyses of nuclear microsatellite DNA loci. Beginning in 2014, we commenced next-generation sequencing (NGS) comparative analyses of the entire mitochondrial genome (mitogenome) of Arctic cod and three other Arctic-adapted gadid species (Polar cod, Walleye pollock, and Saffron cod). In 2015, we commenced NGS RNA-Seq analyses to generate a transcriptome for analysis of functional genes in Arctic cod. Also in 2015, we commenced RAD-Seq analysis of the Arctic cod genome for selection of neutral and functional genomic sequences that could be used for species identification. Data obtained during this effort for both microsatellite and mtDNA cytb data, as well as transcriptomic and mitogenomic data are publicly available in GenBank (https://www.ncbi.nlm.nih.gov/) and in Wilson et al. (2018).

1.5 References

Bradstreet, M.S.W., K. J. Finley, A. D. Sekarak, W. B. Griffiths, C. R. Evans, M. F. Fabjan, and H. E. Stallard. 1986. Aspects of the biology of Arctic cod (*Boreogadus saida*) and its importance in Arctic marine food chains. Canadian Technical Report of Fisheries and Aquatic Sciences No. 1491. 193 pages.

Chen, L., A.L. DeVries, and C.-H.C. Cheng. 1997. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. Proceedings of the National Academy of Sciences USA 94:3817–3822.

Cheung, W.W., V.W. Lam, and D., Pauly. 2010. Bioclimate Envelope Model to Predict Climate- Induced Changes in Distribution of Marine Fishes and Invertebrates, in Cheung, W.W.L., V.W.Y., and Pauly, D., eds., Modeling present and climate-shifted distribution of marine fishes and invertebrates: Vancouver, Univerisyt of British Columnia, Fisheries Centre Reseasrch Reports, v. 16, no. 3, p. 5-50.

Crawford, R. E., and J. K Jorgenson. 1996. Quantitative studies of Arctic cod, *Boreogadus saida* schools: Important energy stores in the Arctic food web. Arctic 49:181–193.

Drost, H. E., M. Lo, E.C. Carmack, and A.P. Farrell. 2016. Acclimation potential of Arctic cod (*Boreogadus saida* Lepechin) from the rapidly warming Arctic Ocean. Journal of Experimental Biology: doi: 10.1242/jeb.140194.

Hughes, A. R., B. D. Inouye, M. T. J. Johnson, N. Underwoood, and M. Vellend. 2008. Ecological consequences of genetic diversity. Ecology Letters 11:609-623. doi:10.1111/j.1461-0248.2008.01179.x

Laurel, B.J., M. Spencer, P. Iseri, and L.A. Copeman. 2016. Temperature-dependent growth and behavior of juvenile Arctic cod (*Boreogadus saida*) and co-occurring North Pacific gadids. Polar Biology 39:1127–1135.

Markert, J.A., D.M. Champlin, R. Gutjahr-Gobell, J.S. Grear, A. Kuhn, T.J. McGreevy Jr., A. Roth, M.J. Bagley, and D.E. Nacci. 2012. Population genetic diversity and fitness in multiple environments. BMC Evolutionary Biology 10:205.

McNeely, J.A., K.R. Miller, W.V. Reid, R.A. Mittermeier, and T.B. Werner. 1990. Conserving the world's biological diversity. World Conservation Union, World Resources Institute, Conservation International, World Wildlife Fund-US, and the World Bank, Washington, D.C.

Mueter, F.J., J. Nahrgang, R., John Nelson, and J. Berge. 2016. The ecology of gadid fishes in the circumpolar Arctic with a special emphasis on the polar cod (*Boregadus saida*). Polar Biology 39:961. https://doi.org/10.1007/s00300-016-1965-3

Nahrgang, J., Ø. Varpe, E. Korshunova, S. Murzina, I. G. Hallanger, I. Vieweg, and J. Berge. 2014. Gender specific reproductive strategies of an Arctic key species (*Boreogadus saida*) and implications of climate change. PLoS One 9:e98452.

Norcross, B.L., S.J., Apsens, Bell LE, Bluhm BA, Dissen JN, Edenfield LE, Frothingham A, Gray BP, Hardy SM, Holladay BA, Hopcroft RR, Iken KB, Smoot CA, Walker KL, Wood ED. 2016. US-Canada Transboundary Fish and Lower Trophic Communities: Abundance, Distribution, Habitat and Community Analysis. Draft Final Report v.2 for BOEM Agreement Number M12AC00011. University of Alaska Fairbanks. 447 pp + appendices.

O'Brien, S.J. and J.F. Evermann. 1988. Interactive influence of infectious disease and genetic diversity in natural populations. Trends in Ecology and Evolution 3:254–259.

Page, L. M., H. Espinosa-Pérez, L. T. Findley, C. R. Gilbert, R. N. Lea, N. E. Mandrak, R. L. Mayden, and J. S. Nelson. 2013. Common and scientific names of fishes from the United States, Canada, and Mexico, 7th edition. American Fisheries Society, Special Publication 34, Bethesda, Maryland.

Pálsson, S., T. Källman, J. Paulsen, and E. Árnason. 2009. An assessment of mtDNA variation in Arctic gadoids. Polar Biology 32:471–479.

Pörtner, H.O. and R.C. Playle (eds). 1998. *Cold Ocean Physiology*. Society for Experimental Biology Seminar Series, Vol 66. Cambridge University Press, Cambridge.

Reed, D.H. and R. Frankham. 2003. Correlation between fitness and genetic diversity. Conservation Biology 17:230–237.

Renaud, P.E., J. Berge, Varpe, O.J. Lonne, J. Nahrgang, C. Ottesen, and I. Hallanger. 2012. Is the poleward expansion by Atlantic cod and haddock threatening native polar cod, *Boreogadus saida*? Polar Biology 35:401–412.

Talbot, S.T., G.K. Sage, S.A. Sonsthagen, and M.C. Fowler. 2014. Arctic cod pilot genetics study: preliminary results from analyses of mitochondrial DNA. U.S. Dept. of the Interior, Bureau of Ocean Energy Management, Alaska BOEM region Anchorage, Alaska. OCS Study BOEM 2014-050. 19 pp.

Thorsteinson, L.K., and Love, M.S., eds. 2016. Alaska Arctic marine fish ecology catalog: U.S. Geological Survey Scientific Investigations Report 2016-5038 (OCS Study, BOEM 2016-048), 768 p., http://dx.doi.org/10.3133/sir20165038.

Vandewoestijne, S., N. Schtickzelle, and M. Baguette. 2008. Positive correlation between genetic diversity and fitness in a large, well-connected metapopulation. BMC Biology 6:46.

Welch, H.E., M.A. Bergmann, T.D. Siferd, K.A. Martin, M.F. Curtis, R.E. Crawford, R.J. Conover, and H. Hop. 1992. Energy flow through the marine ecosystem of the Lancaster Sound region, Arctic Canada. Arctic 45:343–357.

Wildes, S.L., J. Whittle, H. Nguyen, and J. Guyon. 2016. *Boreogadus saida* genetics in the Alaskan Arctic. US Dept. of the Interior, Bureau of Ocean Energy Management, Alaska OCS Region. OCS Study BOEM 2011-AK-11-08 a/b. 67 pp. – DRAFT REPORT.

Wilson, R.E., B.J. Pierson, G.K. Sage, M.C. Gravley, D.M. Menning. S.A. Sonsthagen, J. Nelson, and S.L. Talbot. 2018. Genetic and genomic data on Arctic cod (*Boreogadus saida*) and other gadids, Beaufort Sea, Alaska, 2011-2014. U.S. Geological Survey data release. http://doi.org/10.5066/F7CF9P23.

2 Population Genetic Structure and Diversity of Arctic Cod

2.1 Introduction

Generally, marine fish are thought to exhibit little or no genetic population structure, likely due to lack of geographic barriers across large areas and high dispersal and migration capabilities of pelagic larvae and adults, respectively. Recently, however, numerous studies have shown that many marine fish exhibit significant genetic structure at both macro- and micro-geographic scales (e.g. Salmenkova. 2011, Sá-Pinto et al. 2012, Shum et al. 2014, Karlsen et al. 2014). Previous genetic analyses of Arctic cod that focused on populations in the northern Atlantic Ocean (Fevolden et al. 1999, Pálsson et al. 2008) and the Chukchi Sea (Wildes et al. 2016) uncovered little population differentiation within these areas. However, a recent study of Arctic cod using microsatellite data uncovered genetic structuring associated with fjord and shelf habitats from the North Atlantic Ocean (Madsen et al. 2016). This finding is similar to what has been observed in Atlantic cod (Karlsen et al. 2014, Pampoulie et al. 2015), where local adaptation may have played a role in driving genetic divergence.

Environments with high energy demands, such as the Arctic, may be associated with specific bioenergetics leading to genetics-based metabolic adaptation. In addition, temperature has a large influence on shaping marine ecosystems, in which species, particularly ectotherms, are often adapted to narrow environmental temperatures. For example, fish occupying high latitude ecosystems typically display narrower thermal tolerances than their lower-latitude counterparts, suggesting that fishes in the Arctic may be more sensitive to changes in temperature. Given the hypothesized sensitivity of Arctic fishes to increased temperature change, it is reasonable to suspect that the Arctic cod may have difficulty acclimating to warming temperatures already observed in the Beaufort and Chukchi seas (Laurel et al. 2016). The mitochondrial genome has been shown to be under strong positive selection (Fonseca et al. 2008, Foote et al. 2010, Scott et al. 2010, Garvin et al. 2011), often associated with thermal adaptation and aerobic capacity (Galtier et al. 1999), as well as potentially driving the evolution of the nuclear genome (Rand et al. 2004, Lane and Martin 2010, Burton et al. 2013, Hill 2015). As such, proteins within the mitogenome of Arctic species may be important mediators of physiological processes associated with bioenergetics in high latitude ecosystems (Amo and Brand 2007, Correa et al. 2012).

Within Alaskan waters, the physical oceanography of the Beaufort Shelf is influenced by a variety of local and remote processes including atmosphere, adjacent shelves from Chukchi Sea (western boundary) and Canadian Beaufort Sea shelf (eastern boundary), Arctic Ocean basin, and freshwater discharge from coastal rivers (MBC Applied Environmental Sciences 2003) which may act as dispersal barriers. In addition, the effects of climate change are more pronounced in the Chukchi and Beaufort seas which were historically characterized by year-round ice on the continental shelf but are now experiencing extensive sea ice loss during the summer (Comiso et al. 2008, Stabeno et al. 2012, Wang et al. 2012, Howell et al. 2016) and potentially influencing the distribution and population structure of Arctic Cod (Astthorsson 2016). To explore the population genetic structure of Arctic cod in Alaska waters, we analyzed fragment data from nuclear microsatellite DNA loci and nucleotide sequence data from the mtDNA cytb gene to investigate the population genetic structure of Arctic cod between and within the Alaskan Beaufort and Chukchi seas and considered levels of genetic partitioning was associated with near-shore or off-shore habitats. To determine if there is any major regional biogeography breaks within Alaskan marine waters and to determine if selection imposed by such factors as temperature was acting on the mitogenome, we used whole mitogenome sequence data on a subset of samples used in the Arctic cod population genetics analyses, and three other cod species: Saffron cod, Polar cod, and Walleye pollock.

As the Arctic ecosystem continues to change, distributions will likely shift altering species interactions and potentially species abdunace in areas. Changes in distributions for these two species may result in

new areas of secondary contact and potentially hybridization. Although hybridization is often viewed as having negative consequences on biodiversity, as it may break up co-adapted gene complexes for local adaptations (Barton and Hewitt 1989), the effects of introgressive hybridization (the incorporation of alleles from one species into the gene pool of a second species via hybridization) can be beneficial, and hybridization may play a major role in the generation of genetic diversity in a species' response to changing enviroments (Templeton 1989, Dowling et al. 2016). Species that frequently exchange genes are often able to capture local adaptations via positive selection and maintain ecological, morphological and genetic integrity in the face of high frequency of hybridization. Such interspecies dynamics provide a mechanism for species to increase in genetic diversity that in turn may facilitate rapid species response to environmental changes. In fact, there is genomic evidence that adaptation to recently colonized environments was facilitated by hybridization between modern and archaic ancentral humans and positive selection on certain genes that provided an important reservoir of advantageous alleles in modern humans (Gittleman et al. 2016). Understanding the frequency of hybridization in Arctic cod, if it occurs at all, and its potential impact on species is likely hampered by the fact that the identification of Arctic and Polar cod via external features is difficult (Bouchard et al. 2016). In light of potential changes to species interactions in response to changing environmental conditions and taxonomic identification, a suite of genetic markers is needed to aid in accurate identification of species, monitor changes in species distributions, and quantify the frequency of hybridization. We used ddRAD-Seq protocol to scan the genome of Arctic and Polar cod to identify loci with high degree of divergence that can be used to verify field identifications.

2.2 Methods: Population structure within Chukchi and Beaufort Sea

2.2.1 Sample Collection and DNA Extraction

We were provided fin clips from Arctic cod in the southern Alaskan Beaufort Sea (n = 780) and Chukchi Sea (n = 85) during 2008, 2011, and 2013-2015 and from the Gulf of St. Lawrence, Nova Scotia, Canada from 2010 (n = 30) (Figure 2-1). We stored fin clips in ethanol and muscle in tissue preservation buffer. BOEM Central Beaufort Sea Survey and US/Canada Transboundary Survey conducted sampling in the Alaskan Beaufort along latitudinal transects designed to cover the entire continental shelf in the western, central, and eastern Alaskan Beaufort Sea from Point Barrow to the Alaska-Canadian Boundary, where the continental shelf is very narrow and contains numerous submarine valleys. These transects sampled the inner shelf area (depth < 50m) including the coastal zone (depth < 20m) and the shelfbreak zone (depth > 50m) including the continental slope. We broadly defined regions within the Beaufort Sea as: 1) western, 2) central, and 3) southern (Figure 2-1). The southern Beaufort was further segregated into Camden Bay and southern shelf areas. Within the Chukchi Sea, samples were collected from two general areas: 1) southern (approx. 67°N -169°W) and 2) eastern (approx. 72°N -164°W) Chukchi Sea. Exact locality information for samples collected in 2008 is not known and therefore these samples were only used in analyses at the regional (e.g. sea) level. Detailed sample information is available in Wilson et al. (2018). We extracted genomic DNA from samples collected prior to 2015 using a "salting out" procedure described by Medrano et al. (1990) with modifications described in Sonsthagen et al. (2004) and, as needed for low quality samples, Talbot et al. (2011). We extracted samples from the Chukchi Sea collected in 2015 using QIAGEN DNeasy Tissue Kit (QIAGEN, Valencia, CA) following the manufacturer's recommendations. We quantified genomic DNA concentrations using fluorometry and diluted to 50 ng mL^{-1} working solutions.

2.2.2 Laboratory Techniques – DNA Sequencing and Microsatellite Genotyping

We amplified the mtDNA cytb gene using primers designed using a reference mitogenome available on GenBank (Accession no. AM919428). The resulting primer pairs amplified either a 818 bp (base pairs)

portion of cytb or a 1,266 bp fragment encompassing the entire gene and portion of the adjacent tRNA-Glu and tRNA-Thr (Table S-1). We conducted polymerase chain reaction (PCR) amplifications in a 20 μL volume with 2–100 ng genomic DNA, 0.5 μM each primer, 1.0 μM dNTPs, 10× PCR Perkin Elmer Cetus #1 buffer, and 0.2 units Taq polymerase. PCR reactions began with 94 ºC for 10 min followed by 40 cycles of 94°C for 30 sec, 50°C for 30 sec, and 72°C for 90 sec with a 30 min final extension at 72°C. For the shorter fragment, we excluded the final 30 min extension. We used ExoSAP-IT® (USB Corporation, Cleveland, OH) to remove excess primers and dNTPs in PCR products. We cycle-sequenced all samples using the shorter fragment primer pair which yielded a 707 bp of sequence product consistent across all individuals on either a LI-COR 4200LR or ABI 3730xl. We then reconciled sequences from opposite strands using LI-COR eSeq imaging software and aligned using AlignIR 2.0 or Sequencher 4.1.2 (Gene Codes Corporation, Ann Arbor, MI). We list haplotype assignments and GenBank Accession numbers for each fish sequenced in Wilson et al. (2018).

We amplified eleven microsatellite loci in four multiplexed PCR reactions (BSA6, BSA7, BSA14, BSA15, BSA60, BSA101, Nelson et al. 2013; GMO8 and GMO34, Miller et al. 2000; PGmo32, Jakobsdottir et al. 2006; TCH14, O'Reilly et al. 2000; PGmo127, Skirnisdottir et al. 2008). We chose these loci for analysis to facilitate eventual pooling of fragment data as part of collaborative research with John Nelson (U. Victoria, Victoria, British Columbia, Canada), which seeks to understand population genetic structure in Arctic cod across the North American Arctic. Six primer pairs were redesigned from the published primer sequences to shorten the length of amplified fragment (Table S-1) to reduce the likelihood of allelic dropout in low quality samples (e.g., Sefc et al. 2003). PCR amplification and electrophoresis followed protocols described in Sonsthagen et al. (2004). We determined genotypes for each individual using Gene Profiler 4.05 (Scanalytics, Inc.). For quality control purposes, we amplified and genotyped in duplicate ten percent of the samples for the 11 microsatellite loci. We assessed level of error due to allelic dropout, null alleles, or scoring error for microsatellite loci using MICROCHECKER (van Oosterhout et al. 2004). We note that 8 individuals possessed alleles in the range observed only in Polar cod for microsatellite locus Gmo8 (Madsen et al. 2009, Nelson et al. 2013). Although a single locus genotype is not sufficient to conclude that these individuals are Polar cod, we removed data for these eight individuals from all subsequent analyses. We sampled all presumed Polar cod samples from the southern shelf and Camden Bay locales in the southern Beaufort Sea. We list microsatellite genotype data, including the eight individuals assumed to be Polar cod, in Wilson et al. (2018).

2.2.3 Genetic Diversity and Population Subdivision – Among Seas

We calculated basic population genetic parameters, haplotype (h) and nucleotide (π) diversity, for mtDNA cytb using ARLEQUIN ver. 3.5.2.2 (Exoffier and Lischer 2010). We calculated haplotypic richness in Contrib ver. 1.4 (Petit et al. 1998). Due to low sample size for Chukchi Sea samples with exact locality in the mtDNA dataset, we treated the Chukchi Sea as a single sampling area. We constructed an unrooted phylogenetic tree for mtDNA cytb in NETWORK 4.6.1.3 (Fluxus Technology, Clare, United Kingdom) using the median joining network method (Bandelt et al. 1999). For microsatellites, we calculated Hardy-Weinberg equilibrium (HWE) and observed and expected heterozygosities (H_O and H_E, respectively) in GENEPOP'007 (Raymond and Rousset 1995, Rousett 2008) and linkage disequilibrium (LD) and allelic richness for each microsatellite locus and population in FSTAT ver. 2.9.3 (Goudet 1995); we determined significance following application of Bonferoni corrections for multiple tests. We assessed the degree of population subdivision among seas in ARLEQUIN by calculating pairwise F_{ST} and Φ_{ST} (Excoffier et al. 1992) for mtDNA data and F_{ST} (θ, Weir and Cockerham 1984) and R_{ST} (Slatkin 1995) for microsatellite data. Because samples sizes varied among designated populations (Goudet et al. 1996), we also determined population differentiation based on χ^2 distributions of alleles/haplotypes (genic differentiation) using GENEPOP'007. We adjusted critical values ($\alpha = 0.05$) for F_{ST}, R_{ST} and χ^2 genic differentiation tests for microsatellite data for multiple comparisons using Bonferoni corrections.

To further explore population structure among seas, we used a Bayesian-clustering program, STRUCTURE 2.2.3 (Pritchard et al. 2000) using the autosomal microsatellite data set using the LOCPRIOR option. This model can to detect population structure in datasets with a weak signal of structure not detectable under standard models (Hubisz et al. 2009). STRUCTURE assigns individuals to populations maximizing Hardy-Weinberg equilibrium and minimizing linkage disequilibrium. We ran the analysis for K = 1 – 6, where K is the number of populations, using an admixture model with 100,000 burn-in iterations and 1,000,000 Markov chain Monte Carlo (MCMC) iterations with 10 independent replicates per K. We used the ΔK method of Evanno et al. (2005) and evaluated the estimate of the posterior probability of the data given K, Ln P(D), to determine the most likely number of groups at the uppermost level of population structure.

2.3 Methods: Comparative Phylogeography – Mitogenome

2.3.1 Sample Collection and DNA Extraction

We obtained fin clips or tissue samples from various locations within Alaskan and western Canadian marine waters (Figure 2-2) for four sympatric cod species: Arctic cod (n = 63), Walleye pollock (n = 29), Polar cod (n = 19) and Saffron cod (n = 57). Detailed sample information is available at Wilson et al. (2018). We extracted genomic DNA using the Qiagen Blood & Tissue Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions and quantified by fluorometry using a Quant-iT dsDNA Broad-Range Kit (Invitrogen, Carlsbad, CA USA).

2.3.2 Mitogenome Sequencing and Annotations

We used Long-Range PCR (LR-PCR) to amplify the mitogenome of each individual in 5-7 fragments with at least 500bp of overlap between amplicons. We designed primers for LR-PCR amplification (Table S-2) by aligning published mitogenomes of Arctic cod (GenBank accession AM919428), Polar cod (AM919429), Greenland cod (*Gadus ogac*; DQ356941), Walleye pollock (AB182307) and Atlantic cod (*Gadus morhua*; NC002081). Each 20 µL PCR reaction for each primer pair contained 2.0µL 10x LA PCR Buffer II (25 mM Mg^{2+}), 3.20 µL 2.5 mM dNTP mixture, 0.1 µL TaKaRa LA T*aq* (Takara Bio Company, Mountain View, CA, USA), 0.8-1.50 µL 10 µM of each primer, 1 µL of template DNA, and 10.60-12.00 µL of RNA-free water. For samples that failed to amplify, we added 0.80 µL of PCR additive 5M Betaine to subsequent PCR reactions. PCR conditions were 94°C for 5 minutes followed by 30 cycles of 98°C for 15 seconds, 50°C for 15 secs and 68°C for 11 minutes, ending with a final extension of 72°C for 10 minutes.

We purified the resulting amplicons for each individual using ExoSAP-IT (Affymetrix, Cleveland, OH USA), quantified with Quant-iT dsDNA High Sensitivity Kit (Invitrogen, Carlsbad, CA USA), and then pooled by individual in equimolar concentrations. We pooled amplicons for each individual sample to construct libraries using a Nextera XT 96 sample DNA kit using standard Illumina barcodes (Illumina, San Diego, CA USA). We performed next-generation sequencing following manufacturer's protocol on a MiSeq Desktop Sequencer (2 × 250 bp read-length configuration).

2.3.3 Mitogenome Assembly and Verification

We recovered each sample by identifying individual barcodes and performed adaptor trimming with MiSeq Reporter software (Illumina, San Diego, CA USA). We assembled and mapped the raw paired-end reads to their respective species reference mitogenome (see above) in Geneious 8.1.8 (BioMatters Limited, Auckland, New Zealand) using high sensitivity setting and minimum mapping quality set to

greater than 99.9% (minimum confidence that read is correctly mapped). Since a Saffron cod reference mitogenome was not publicly available, we reconstructed mitogenomes from four samples with a subset of reads (~7,000), using the bait and iterative mapping approach in the programs MIRA 3.4.1 (Chevreux et al. 1999) and MITOBIM v 1.8 (Hahn et al. 2013) to serve as a reference. This approach has shown to produce high quality mitogenomes even with low-coverage data (Anmarkrud and Lifjeld 2016, Machado et al. 2016).

To detect the presence of pseudogenes or gene arrangements in the consensus mitogenomes, we annotated each mitogenome using MITOannotator (http://mitofish.aori.u-tokyo.ac.jp/; Iwasaki et al. 2013, Wataru et al. 2013). In addition, we conducted Sanger sequencing of a 818 bp fragment of mtDNA cytb gene on samples from 16 Arctic cod, using methods similar to those presented above and primers given in Table S-3. We used those sequences, and sequences previously published as single mtDNA gene sequences for Walleye pollock and Saffron cod, to verify that a nuclear pseudogene was not included in the consensus mitogenomes. We then constructed strict consensus sequences for each sample and aligned using Muscle (Edgar 2004) within Geneious using default settings, with the final alignment inspected visually. We archived annotated genomes in GenBank and detailed information on samples and GenBank accession numbers can be found in Wilson et al. (2018).

2.3.4 Genetic Diversity and Population Structure

We calculated numbers of haplotypes (H), haplotype diversity (Hd) and nucleotide diversity (π) in DnaSP v5.1 (Librado and Rozas 2009) for each of the 13 protein-coding genes, tRNAs, rRNAs, control region and over the entire mitogenome for each species.

To estimate levels of genetic differentiation, we first broadly assigned samples to geographic regions based on sampling locality (Figure 2-2). These regions included southern localities: Gulf of Alaska (Saffron cod), Aleutians (Walleye pollock), Bering Sea (Walleye pollock, Arctic cod, and Saffron cod) and regions north of the Bering Strait: Chukchi Sea (Walleye pollock, Arctic cod, and Saffron cod), and Beaufort Sea (Arctic cod). Since only one locality was sampled for Polar cod, no inter-population analysis could be performed. We calculated F_{ST} for the 13 protein-coding and two rRNA genes separately in ARLEQUIN v3.5.2.2 (Excoffier and Lischer 2010). We also used a Bayesian clustering method in Splitstree ver 4.14.2 (Huson and Bryant 2006) to compute phylogenetic networks. We computed Neighbor-Net networks (Bryant and Moulton 2004) based on genetic distances (we used *P*-distances) among individuals to visualize genetic variation within and among regions.

2.3.5 Demographic History, Tests of Neutrality and Analyses of Selection on Mitochondrial DNA

We calculated Tajima's *D* (Tajima 1989) and Fu's *Fs* (Fu 1997) in ARLEQUIN to distinguish between neutral evolution of each of protein-coding genes (e.g. genetic drift), non-random processes (e.g. selection) and demographic expansion or contraction. Tajima's *D* uses the frequency of segregating nucleotide sites while Fu's *Fs* uses the distribution of alleles. Significant negative values could indicate purifying selection while significant positive values may indicate balancing selection. Both statistics are also sensitive to demographic processes where a recent expansion can also leave a low or negative test statistic.

To generate the required phylogeny needed for codon-based selection tests, we used a Bayesian phylogenetic analysis implemented in MrBayes 3.2.5 (Ronquist et al. 2012) on a concatenated alignment of the 13 protein-coding genes. Gaps, intergenic spacers, and stop codons were removed. Since the gene ND6 is encoded on the light strand, we used the reverse-complemented sequence for phylogenetic reconstruction. In addition, we repeated overlapping bases of ATPase8/ATPase6 and ND4L/ND4 in the alignment to enable analyses on the all codons. We determined the best-fit partitioning scheme and

nucleotide substitution models using a heuristic search (greedy algorithm) under the Bayesian information criterion (BIC) in PartitionFinder (Lanfear et al. 2012; see Table S-4). We ran four replicate runs with chains for 2 million generations sampled every 1000 generations. Burbot (*Lota lota*, GenBank accession NC004379) was used as an outgroup based on phylogeny reconstruction by Roa-Varón and Ortí (2009). We assessed run convergence using TRACER 1.6.0 (Rambaut et al. 2014) and found that the Potential Scale Reduction Factor (PSRF) was close to 1.0 for all parameters. We summarized trees as a consensus tree (Figure 2-3A) using a 50% majority rule following a 25% burn-in and visualized with FIGTREE 1.4.2 (Rambaut 2009). We used the consensus tree with outgroup (*Lota lota*) removed (Figure 2-3B) in the codon-based selection analyses below.

We used DnaSP to calculate McDonald–Kreitman tests of directional or purifying selection (MK; McDonald and Kreitman 1991) for each gene and complete mitogenome for each species. The MK test uses the ratios of non-synonymous and synonymous fixed differences between species and polymorphic differences within species. A Neutrality Index (NI) greater than 1.0 indicates purifying selection and a NI less than one was interpreted as positive selection.

We used the Fast Unconstrained Bayesian AppRoximation (Murrell et al. 2013) and Mixed Effects Model of Evolution (MEME; Murrell et al. 2012) in the program HyPhy, implemented on the web-server Datamonkey (Kosakovsky Pond et al. 2005, Delport et al. 2010), to identify codons putatively under selection. The Fast Unconstrained Bayesian AppRoximation assigns each codon a posterior probability (PP) of belonging to three classes of dN/dS (ω): $\omega < 1$, $\omega 1 = 1$ and $\omega > 1$. We inferred codons with PP > 0.9 and $\omega > 1$ or $\omega < 1$ to have evolved under positive and purifying selection, respectively. While the Fast Unconstrained Bayesian AppRoximation identifies codons putatively under selection across all branches on the phylogeny, MEME allows ω at each codon to vary across branches/lineages and so detects episodic selection. We considered codons with $P < 0.05$ to have experienced episodic positive selection.

We used two approaches to evaluate the potential changes in physiochemical properties associated amino acid replacements across the phylogeny: 1) TREESAAP and 2) PRoperty Informed Models of Evolution (PRIME). TREESAAP v3.2 (Woolley et al. 2003) relies on the MM01 model implemented in BASEML (Yang 1997). We considered only amino-acid changes with radical effects (categories 6, 7, 8) and z-scores above 3.09 ($P < 0.001$) as these effects are likely to change the protein biochemistry and are therefore potentially under positive selection (McClellan et al. 2005). We used a global test to select properties under selection across all coding regions and used a sliding-window approach (window size of 10 codons with 1-codon increments) to detect regions under selection.

We also used PRIME on the web-server Datamonkey to characterize the potential change in properties of the residues that experience positive dN/dS skew. PRIME is a variation of MEME that detects residue changes that can be categorized as changes in the original property of the amino acid. There are five categorical changes that can be detected using PRIME: polarity index, secondary structure factor, volume, refractivity/heat capacity and charge/iso-electric point (Atchley et al. 2005, Conant et al. 2007). PRIME was run using both sets of amino acid properties. A change in these properties was considered significant if the posterior probability was greater than 0.9. To reduce the impacts of false inferences on selection, we only considered amino acid sites that were supported by at least two methods and were fixed within species (i.e. not present in tip lineage) to be potential candidates to be under positive selection.

2.4 Methods: ddRAD Sequencing – Species Identification

2.4.1 Sample Collection and DNA Extraction

We extracted genomic DNA from fin clips or tissue of Arctic cod (n = 123) and Polar cod (n = 14), the latter identified morphologically using mouth structures (personal communication Andrew Majewski, Fisheries and Oceans Canada), using a DNeasy Blood & Tissue kit and following the manufacturer's protocols (Qiagen, Valencia, CA, USA). Where possible, we used samples from which mtDNA cytb data were available. We quantified extractions using a Modulus Microplate (Turner BioSystems, Inc.) and a Broad Range Quant-iT dsDNA Assay Kit (Thermo Fisher Scientific, Inc.) to ensure a minimum concentration of 10 ng/μL. We included one sample (16973, collected during 2014) that, based on microsatellite data may have been incorrectly identified to species in the field, in the study to verify taxonomic status.

2.4.2 ddRAD-seq Library Preparation

Sample preparation for ddRAD sequencing followed the double-digest protocol outlined in DaCosta and Sorenson (2014). We digested genomic DNA (~1 μg) with high fidelity versions of *Sbf*I and *Msp*I restriction enzymes (New England Biolabs, Ipswich, MA, USA). We ligated amplification and sequencing adapters containing unique barcode or index sequences to the sticky ends generated by the restriction enzymes. We then electrophoresed samples on 2% low-melt agarose gel and selected DNA fragments of size between 300-450 bp (178-328 bp excluding adapters). We extracted DNA from the gel using a MinElute Gel Extraction Kit (Qiagen) following the manufacturer's protocol. We then amplified size-selected fragments via PCR, using Phusion high-fidelity DNA polymerase (Thermo Scientific, Pittsburgh, PA, USA) for 24 cycles, and cleaned the amplified products using magnetic AMPure XP beads (Beckman Coulter, Inc., Indianapolis, IN, USA). We used quantitative PCR via an Illumina library quantification kit (KAPA Biosystems, Wilmington, MA, USA) to quantify the concentration of purified PCR products, and samples were pooled in equimolar concentrations. We sequenced a multiplexed library as a single-end 150-base pair run on an Illumina HiSeq 2500.

2.4.3 Bioinformatics

We processed raw Illumina reads using a computational pipeline described by DaCosta and Sorenson (2014; http://github.com/BU-RAD-seq/ddRAD-seq-Pipeline). We first assigned reads to individual samples based on barcode/index sequences using bcl2fastq-1.8.4 software (Illumina, San Diego, CA). We conducted pre-processing of reads using a custom Python script (developed by Jeff DaCosta, Boston College University) that removed chimera sequences (*Sbf*I-SbfI or *Msp*I-*Msp*I) or reads containing 2 or more mismatches in *Sbf*I recognition sites and trimmed reads to either *Msp*I concatemer or P2 adapter. We then collapsed reads per sample into identical clusters using the *CondenseSequences.py* script with low-quality reads (i.e. sequences that failed to cluster with any other reads (−id setting of 0.90) and an average per-base Phred score < 20) filtered out using the *FilterSequences.py* script and the UCLUST function in USEARCH v.5 (Edgar 2010). With condensed and filtered reads from all samples, we then concatenated and clustered with an −id setting of 0.85, using UCLUST. We used the program MUSCLE v.3 (Edgar 2004) to align and cluster reads, and samples within each aligned cluster were genotyped using the *RADGenotypes.py* script. We identified homozygotes and heterozygotes based on thresholds outlined in DaCosta and Sorenson (2014), with individual genotypes falling into three categories: 'missing' (no data), 'good' (unambiguously genotyped) and 'flagged' (recovered heterozygous genotype, but with haplotype counts outside of acceptable thresholds or with > 2 alleles detected). We retained polymorphic loci with a median depth of 10, < 10% missing genotypes, and < 10% flagged genotypes for downstream analyses,

based on the Polar cod sample size. We skipped loci flagged for manual editing for preliminary analysis, but included all loci in the final dataset.

2.4.4 Level of Divergence

We calculated pairwise Φ_{ST} as well as nucleotide diversity for each ddRAD-seq loci and overall using the R package PopGenome (Pfeifer et al. 2014). We assessed genetic structure by plotting the first two principal components in a principal component analysis (PCA) as implemented in the adegenet R program (i.e. dudi.pca; Dray and Dufour 2007, Jombart 2008).

2.5 Results: Population Structure within Chukchi and Beaufort Seas

2.5.1 Effects of Sampling Time Periods

Genetic samples for mtDNA within the Beaufort Sea were sampled across a seven-year span (2008, 2011, 2013, and 2014). We relied on results of an exact test of population differentiation in ARLEQUIN ver. 3.5.2.2 to determine if there was a correlation of the distribution of haplotypes with year sampled. We found no significant differences in the distribution of haplotypes among sampling years (for each test, $P > 0.05$). To further investigate the relationships among haplotypes across years, we constructed a temporal statistical parsimony network using the R script TempNet v1.4. (Prost and Anderson, 2011; Figure 2-4). We found the same high frequency mtDNA haplotypes (Hap1, Hap2, Hap4, Hap7, and Hap9) within each year's sampling period. For Chukchi Sea samples, we obtained six mtDNA cytb sequences from 2015 and 30 from 2008. We found no significant differences in haplotype distribution between sampling periods ($P = 0.73$). Due to low sample size in 2015 and unknown exact locality within Chukchi Sea of 2008 samples, we treated Chukchi Sea sequences as a single population and only used these sequences for analyses at the sea level (Beaufort Sea vs. Chukchi Sea).

For microsatellites data, Beaufort Sea samples were collected across a four-year time frame (2011-2014) and Chukchi Sea samples were collected from 2008 and 2015. We found no significant differences in the allelic distribution between years ($P = 0.57$). Thus in sea level analyses, 2008 and 2015 were pooled as a single Chukchi Sea population.

2.5.3 Genetic Diversity and Population Subdivision – Among Seas

We observed 82 unique mtDNA cytb haplotypes (Figure 2-5) characterized by 64 variable sites among Arctic cod ($n = 407$) sampled from the Alaskan Beaufort and Chukchi seas and the Gulf of St. Lawrence in Canada. Fifty-six of the haplotypes (68%) were represented by a single individual (i.e. private) with all seven higher frequency haplotypes shared among seas and formed two distinct groups (Figure 2-5). Diversity measures were moderate to high and similar across all sampling regions (Table 2-2).

We obtained microsatellite genotypes from 740 individuals from the Beaufort and Chukchi seas. The average number of alleles per sampling area ranged from 6.2-11.6 alleles per locus with allelic richness values similar across all areas (Table 2-2). The sampling transects in the southern Beaufort Sea (Camden Bay and Southern Shelf) contained the highest number of private alleles (9 and 20, respectively). Overall, heterozygosity levels were similar across sampled locales, and all locales, except the eastern Chukchi Sea ($\chi^2=\infty$, d.f. = 22, $P < 0.0001$), were in HWE (Table 2-2). Two loci, Bsa101 and PGmo32, were in linkage disequilibrium when the analysis included all locales ($\alpha = 0.05$), however when loci pairs were analyzed by each sampled location all loci were in linkage equilibrium. We found no evidence of scoring error due to stuttering or large allelic drop-out in any of the populations, based on MICROCHECKER. Evidence of potential null alleles was detected in southern Chukchi Sea (locus Gmo34) and eastern Chukchi Sea (loci Pgmo32, Gmo8, and Gmo34); this is an unsurprising result since these loci were out of HWE due to

homozygote excess in the respective populations (Table 2-3). Since those loci did not show heterozygote deficit in other populations, we assumed that the homozygote excess found by MICROCHECKER is not due to null alleles but rather to lack of mutation-drift equilibrium in particular in the eastern Chukchi Sea population. Therefore, all loci were retained in further analyses.

We observed no significant differentiation across sampling areas for either mtDNA (overall F_{ST} = -0.001 or Φ_{ST} = -0.010) or microsatellites (overall F_{ST} = 0.001 or R_{ST} < 0.001). In addition, no pair-wise comparison was significant for either the nuclear or mtDNA datasets (Table 2-4). However, we did find a significant difference based on the χ^2 test in the distribution of microsatellite alleles between southern Chukchi Sea area and the central and southern Beaufort Sea sampling localities (Table 2-4). In accordance with the overall differentiation test, Bayesian analysis of population structure (STRUCTURE analysis) revealed little genetic structure across the Beaufort and Chukchi seas. STRUCTURE analysis also indicated a lack of structure between coastal and shelf break stations (Figure 2-6); most likely number of clusters (K) in the dataset was 2, based on Evanno's method (ΔK = 56.9) and the likelihood given the data (LnPR(K=1) -20314.3 vs LnPR(K=2) = -20078.1). The majority of samples were assigned to Cluster 1 (99.3%), with only 4 samples from the Chukchi Sea assigned to Cluster 2 with high probability (> 96%) and one sample from southern Beaufort Sea assigned to Cluster 2 with 45% probability.

2.6 Results: Comparative Phylogeography within Alaskan waters

2.6.1 Mitogenome Assembly and Mapping

Mitogenome assemblies rendered good coverage when mapped against reference [Arctic cod (mean coverage = 5,367 X), Saffron cod (mean coverage = 3,805 X), and Walleye pollock (mean coverage = 5,427.5X)] and showed no evidence of gene rearrangements compared to published reference mitogenomes. In addition, we recovered the duplication in the oriL and flanking tRNA discovered by Breines et al. (2008) that we verified by visual inspection of reads. Six Walleye pollock samples did not amplify PCR fragment 1 (bps 111-3944 in alignment) and therefore we only obtained partial mitogenomes for those samples with mean coverage of the remaining mitogenome of 1,783X. All consensus sequences for Arctic cod, Walleye pollock, and Saffron cod matched previously published sequences or newly obtained Sanger sequenced cytb (see Sections 2.3.1.2 and 2.3.2.3).

For Polar cod, we obtained full or partial mitogenomes for 19 samples (mean coverage = 5,141X). However, the Sanger sequence derived nucleotide sequences of cytb (818 bp) did not match the resulting consensus sequences (92.3% similarity) nor did it match (94.9 % identity) GenBank reference (AM919428). The overall consensus sequences did match the GenBank reference for base pairs 1-13,149 (PCR fragments 1-4) in the Polar cod alignment. As we could not determine if this was a potential cause of mtDNA recombination as observed in other vertebrate species (Mjelle et al. 2008, Pilgrim et al. 2008, White et al. 2008), and to eliminate the impact of potential pseudogenes, we excluded from analyses the base pairs starting at position 13,150 (located in gene ND5) from any analyses.

2.6.2 Genetic Diversity and Population Divergence

No significant population structure (Φ_{ST}) was detected across the entire mitogenome or at each protein-coding gene or rRNAs (Ps > 0.05); the only exception was the Saffron cod Gulf of Alaska population, which was significantly differentiated (P < 0.05 for all tests) for 11 of the 13 protein coding genes from both two northern sampling locales (Bering and Chukchi seas) and was significantly differentiated in the remaining two genes (COII and ND3) between one of the northern locations. The neighbor-net trees showing relationships among mitogenome haplotypes from general sampling locales within each species are provided (Figures 2-7 – 2-10).

In general, tRNAs showed considerably less variability than the 13 protein-coding genes. Overall levels of nucleotide and haplotype diversity were similar across species, with the Arctic cod showing a slightly higher levels of nucleotide diversity at eight of the 13 protein-coding genes (Figures 2-11 and 2-12). In Arctic cod and Walleye pollock, northern regions in general showed higher overall nucleotide diversity than more southern regions, but Saffron cod had similar levels of genetic diversity across regions. However, at the level of the gene, the northern Saffron cod population showed greater nucleotide diversity for multiple coding genes and tRNAs than the Bering Sea and Gulf of Alaska sampling locations (Table S-8).

2.6.3 Neutrality Test

Within each species, we observed significantly negative values for both Tajima's *D* and Fu's *Fs* for most genes (Table S-5 – S-8). At the intra-species level, there were significant negative values within Arctic cod throughout the mitogenome (Table S-5). With Walleye pollock, only CO2, ND3, and ND6 had a significant negative value for either Tajima's *D* or Fu's *Fs* while values for least six genes were negatively significant in most southern regions (Table S-7). Within Saffron cod, there was no signature of population expansion or selection within the Gulf of Alaska (Table S-8).

2.6.4 McDonald–Kreitman Tests

Results of an MK test suggested that strong purifying selection (NI >1) influenced divergence between cold-adapted (Arctic and Polar cod) and non-cold adapted (Saffron cod and Walleye pollock; Fisher's test Ps < 0.05 for all pair-wise tests). An MK test of Arctic cod vs Polar cod was not significant (Fisher's P = 0.21). MK tests performed on individual genes were significant for COII, ND4L, ND4, ND5, and cytb; primarily Saffron cod vs one or two of the other species. The MK tests in Arctic cod vs Walleye pollock were significant for ND4 and ND5 (Fisher's $P < 0.004$) and Arctic cod vs Polar cod for ND4 (Fisher's $P = 0.005$). Amino acid replacements within each species ranged from six (Polar cod) to 32 (Arctic cod), with 90% of intra-species amino acid changes found in a single individual.

2.6.5 Detecting Selection Using Codon Models.

The Fast Unconstrained Bayesian AppRoximation supported strong pervasive purifying (negative) selection across all mitochondrial genes, with 3597 codons (94.7%) showing purifying selection at posterior probability of > 0.90. No codons showed evidence of pervasive diversifying selection. MEME found evidence of episodic diversifying selection at nine codons ($P < 0.05$). Only two of these nine codons were associated with fixed amino acids; the others were associated with tip lineages (i.e. only observed in a single or few individuals within a species; see Table S-9). Both of these codons were associated with the Arctic-adapted Arctic cod (codon 2 in ND1 and codon 74 in ND2; Table 2-5).

2.6.6 Radical Changes in Physiological Properties

The global TREESAAP analysis detected two physiochemical properties (equilibrium constant [ionization of COOH] and α-helical tendencies) that showed an overall pattern of positive selection (z-score > 3.09, $P < 0.001$). Sliding window sizes of 10 and 20 produced similar results; therefore, we only present results from window size 10 analysis. Sliding window analysis showed the equilibrium constant (ionization of COOH) was affected by amino-acid replacements not present in tip lineages in ND1, ND2, ND4, and cytb and α-helical tendencies by replacements in CO3, ND3, ND4, and ND5 (Table 2-5 and Table S-9 and S-10). PRIME suggested that adaptive evolution may have affected three sites (indicated by negative weight of amino acid properties "volume" and "secondary structure factor", see Table 2-5).

Of the fourteen nonsynonymous codons inferred to be under positive selection that did not affect a single tip lineage, only one amino acid was detected by more than one methodology and thus most likely not a false positive. This codon in ND1 involved a replacement of the amino acid threonine with leucine (Thr→Leu) in Arctic cod relative to the other gadid species assayed (Table 2-5).

2.7 Results: ddRAD sequencing – Species Identification

2.7.1 Loci Recovered

We obtained a total of 138,598,600 raw sequencing reads with a maximum 150 bp length using single-end sequencing on an Illumina HiSeq 2500. The number of reads for each individual ranged from 484,801 to 3,325,291 with an average of 989,990. We recovered 897 ddRAD-seq loci with good alignments resulting in 2,671 bi-allelic SNPs (single nucleotide polymorphisms) and 166 polyallelic sites (> 2 alleles). Data are archived in GenBank and in Wilson et al. (2018).

2.7.2 Genetic divergence between species

The overall Φ_{ST} across the 897 ddRAD-Seq loci ranged from -0.02–1.00 between species with an average of 0.49 (Figure 2-13). Approximately 34% of the loci showed an estimated Φ_{ST} value to zero. Conversely, only 4% (39 loci) had a $\Phi_{ST} > 0.999$ (i.e. fixation between species) with 358 loci (40%) showing extremely high levels of divergence ($\Phi_{ST} > 0.90$). Overall nucleotide diversity was similar between species, 0.1144 for Arctic cod and 0.0943 for Polar cod.

Based on visual inspection of genotypes and PCA (Figure 2-14), few samples appeared to be misidentified. One Arctic cod sample (16973) identified as a potential Polar cod based on microsatellite Gmo8 was confirmed as a Polar cod with the ddRAD-Seq loci. Three samples originally identified as Polar cod had Arctic cod-like genotypes; for two of these samples, we were able to obtain whole mitogenomes to verify species lineage. Sample PC 20011414 possessed an Arctic cod mitogenome, confirming it is likely an Arctic cod. However, PC 20015632 had a Polar cod-like mitogenome with an Arctic cod-like nuclear genome, suggesting this individual is of a hybrid origin.

2.8 Discussion

2.8.1 Population Genetic (Stock) Structure and Diversity

Although we detected the presence of two discrete mitochondrial cytb lineages (Group A and B, Figure 2-5) in Arctic cod occupying Alaskan and Canadian waters, there was no discernable geographic clustering of those two lineages, as assessed using traditional population differentiation metrics (F_{ST}, Φ_{ST}, χ^2 distribution of haplotypes). This is concordant with the findings of Pálsson et al. (2009) across a similar geographic range of samples using sequence data from the the mtDNA cytb gene. We also observed lack of regional population structuring for estimates of F_{ST}, R_{ST}, and $\chi 2$ distribution of alleles for the microsatellite data, across locales assayed in the Chuckchi and Beaufort seas. Bayesian clustering analyses detected two discrete clusters within the microsatellite dataset, but the second cluster was comprised of only four individuals and likely reflects the presence of rare genotypes in those individuals. The comparative species approach, leveraging whole mitogenome sequence data, verified the lack of regional population structure within the Walleye pollock and Saffron cod as well as Arctic cod north of the Aleutian Islands. This cross-species approach further suggests a general lack of barriers to movement of Arctic cod within Alaskan Arctic marine waters and that Arctic cod in this region comprise a single panmictic population.

Arctic cod still appear to have some genetic population structure that is not strictly associated with the boundaries between seas and is consistent with isolation-by-distance. A significant differentiation between the most northeastern (central Beaufort Sea) and most southern (Chukchi Sea) Arctic cod samples was detected based on the χ^2 distribution of microsatellite alleles. Among traditional population differentiation metrics, χ^2 tests for significant differences in the distribution of alleles are the most powerful among traditional population differentiation metrics when there are disparities in sample sizes (Goudet et al. 1996). We suspect that a transitional zone occurs in the southern Beaufort Sea, given the high number of private alleles (alleles only found in that particular locale) there. A similar geographic pattern is apparent in zooplankton species composition with similarities between the Chukchi and western Beaufort seas near Point Barrow reflecting an influence of Pacific-origin waters, but a more typical Arctic fauna community further east within the Beaufort Sea with a (Smoot and Hopcroft 2017). Our ability to determine whether there is a transitional zone, or biogeographic break, between two different subgroups of Arctic cod requires additional sampling across the Mackenzie River and locales to the east.

2.8.2 Selection on the Mitogenome

Our analyses uncovered evidence of purifying (negative) selection (removal of deleterious alleles) at 96% of all codons, supporting the hypothesis of purifying selection as the main evolutionary force affecting Arctic cod mitogenome evolution. This result is congruent with other organisms (Nabholz et al. 2013), including certain cod species (Harrison et al. 2015, Pavlova et al. 2017), as purifying selection is expected to preserve the protein-coding function of mitochondrial genes. Negative selection can eliminate extreme values of a trait (that is, divergent alleles), resulting in stabilizing selection, long thought to be a common mechanism for natural selection in which population means stabilizes on a particular, moderate trait values and favors individuals with intermediate phenotypes (Charlesworth et al. 1982, but see Kingsolver et al. 2001). Over time, stabilizing selection can lead to a decrease in genetic diversity that can ultimately hamper resilence when environmental conditions shift.

However, particular sites within the mitochondrial genome can experience positive selection in response to environmental pressures against background of strong purifying selection. Although we found no codons showing evidence of pervasive diversifying selection (that is, extreme traits are favored over intermediate traits), 9 codons showed a signal of episodic diversifying selection and two functionally relevant fixed amino acid differences in the OXPHOS complex 1 (the ND1 and ND2 genes) in Arctic cod relative to the other species. The amino acid replacements in OXPHOS complex 1 (ND genes) tend to have minor effects on the functional properties of amino acids while changes in complex IV (COX genes) tend to have larger effects on amino acid properties (Zhang and Broughton 2013).

Despite the finding that purifying selection at the mitogenome of Arctic cod may be the main evolutionary force impacting that genome's evolution, Arctic cod possess characteristics that may provide this species with the ability to evolutionarily respond (i.e. change in genetic makeup) to a changing environment. Previous genetic studies of Arctic cod in Alaskan waters have characterized this species as having low genetic diversity at the presumably neutral nuclear and mtDNA markers (Wildes et al. 2016), which may reflect low genetic diversity in functional genes and signal a limited capacity to respond adaptively to changing ocean conditions. However, relative to other co-distributed cod species, Arctic cod possess moderate to high levels of genetic diversity, with higher levels detected in the Beaufort Sea where the effects of climate change are most pronounced (Comiso et al. 2008, Stabeno et al. 2012, Wang et al. 2012, Howell et al. 2016). Haplotype diversity levels are also higher than reported in Wildes et al. (2016). It is likely that the large effective population size allows for the retention of high genetic diversity (Morgan-Richards et al. 2017). In addition to maintaining genetic diversity, selection tends to be more efficient in large populations since mutation can act on more individuals and beneficial mutations can arise independently on different genetic backgrounds (Barton 2010, Ralph and Coop 2010). Although most of the amino acid differences between species are likely the result of purifying selection rather than

positive selection, there is evidence of positive selection acting on certain codons in the mitogenome in particular within the Arctic cod, which experience a narrower thermal habitat preference. These factors — high mtDNA genetic diversity and selection — are potentially important drivers of cod evolution, suggesting that Arctic cod may have the potential to respond at changing marine environments.

2.8.3 Species Identification and Hybridization

The finding of Polar cod-like alleles in Beaufort Sea Arctic cod at a single microsatellite DNA locus raises questions about field identification of cod and identifying appropriate genetic markers to differentiate species and/or detect possible introgressive hybridization among fish species. Currently, a single microsatellite is used to distinguish Arctic from Polar cod (Madsen et al. 2009), although there is some minor overlap in allele sizes. However, a single microsatellite and/or mtDNA is unlikely to be sufficient to differentiate between potential reasons for a mismatch between species determination based on genetic characters and determination based on morphological or distributional characters; that is, the mismatch can be due to simple field misidentification, or the failure to detect a hybrid. While first generation (F1) hybrids should possess the mitogenome of one species and biparentally-inherited alleles from the nuclear genome of both species, decay in later generations can mask the parental contribution of a hybrid's descendants. Such is the case for sample PC 20015632 from the western Canadian Beaufort Sea. This sample was identified as a Polar cod based on general morphology and mouth structures (Andrew Majekski, Fisheries and Ocean Canada, personal communication). Nuclear (ddRAD-Seq) markers confirmed that this sample is a Polar cod. However, mitochondrial analysis indicates this fish is an Arctic cod. It has been shown that species can effectively lose the genetic signature of hybridization within four generations of backcrossing (Lavretsky et al. 2016) suggesting this sample is at least a F5 hybrid. Our finding suggests that Arctic cod x Polar cod hybrids are fertile and therefore capable of reproducing. Therefore, this study highlights the need of using multiple markers with different inheritance (mitochondrial and nuclear markers) for species identification.

Generally, geographical isolation is a major factor in limiting the impacts of hybridization and introgression. However, with current and likely continuing shifts in marine fish distributions, Arctic marine ecosytems are predicted to have the largest shift in species composition (Fossheim et al. 2015, Frainer et al. 2017) due to such processes as competitive exclusion, exposure to new pathogens/parasites, and hybridization (Jørgensen et al. 2016). Hybridization has been proposed to have a large influence on the evolution of cod-fishes and may have resulted in new species that become very common. For example, researchers have proposed that the commercially important Walleye pollock is the result of hybridization between Arctic and Atlantic cod (Holldórsdóttir and Árnason 2015). Further, hybridization can have negative impacts such as reducing fitness (Muhlfeld et al. 2009) or it can increase genetic diversity by creating new combinations of genes needed to adapt to changing environment (Hedrick 2013). It is unclear whether the instance of hybridization observed between Arctic cod and Polar cod is a rare but regular event, or whether it is a relatively new possibility linked to reduction in sea ice that minimize geographic isolation. Unfortunately, the current distribution of Polar cod in this region is not well known (Thorsteinson and Love 2016). Reports have indicated that Polar cod is rare in the U.S. Beaufort Sea, with only three reported specimens, but is abundant near the Mackenzie River outflow in Canada (Thorsteinson and Love 2016). Our detection of eight Polar cod in the eastern U.S. Beaufort doubles the number of reported observations. This suggests Polar cod may be more abundant in the U.S. Beaufort Sea than previously thought, possibly due to misidentification of specimens based on morphology. Furthermore, our finding of a hybrid sample in the western Canadian Beaufort at the edge of the main distribution suggests this region may represent a natural hybrid zone, but hybridization between the two species may also be an infrequent event. Additional understanding of how hybridization has influenced genetic diversity of Arctic cod and the species' propensity to hybridize may provide insight on how Arctic cod in the Beaufort Sea may respond to a changing environment.

2.9 References

Amo, T., and M.D. Brand. 2007. Were inefficient mitochondrial haplogroups selected during migrations of modern humans? A test using modular kinetic analysis of coupling in mitochondria from cybrid cell lines. Biochemical Journal 404:345–351.

Anmarkrud, J.A., and J.T. Lifjeld. 2016. Complete mitochondrial genomes of eleven extinct or possibly extinct bird species. Molecular Ecology Resources doi:10.1111/1755-0998.12600.

Astthorsson, O.S. 2016. Distribution, abundance and biology of polar cod, *Boreogadus saida*, in Iceland–East Greenland waters. Polar Biology 39:995–1003.

Atchley, W.R, J. Zhao, A.D. Fernandes, and T. Drüke. 2005. Solving the protein sequence metric problem. Proceedings of the National Academy of Sciences USA 102:6395–6400.

Bandelt H. J., P. Forster, B. C. Sykes, and M. B. Richards. 1995. Mitochondrial portraits of human populations. Genetics 141:743–753.

Barton N.H., and G.M. Hewitt. 1989. Adaptation, speciation and hybrid zones. Nature 341:497–502.

Barton, N. 2010. Understanding adaptation in large populations. PLoS Genet. 6: e1000987.

Bouchard, C., S. Mollard, K. Suzuki, D. Robert, and L. Fortier. 2016. Contrasting the early life histories of sympatric Arctic gadids *Boreogadus saida* and *Arctogadus galcialis* in the Canadian Beaufort Sea. Polar Biology 39: 1005–1022.

Burton, R.S., R.J. Pereira, and F.S. Barreto. 2013. Cytonuclear genomic interactions and hybrid breakdown. Annual Review of Ecology, Evolution, and Systematics 44:281–302.

Breines, R., A. Ursvik, M. Nymark, S. D. Johansen, and D. H. Coucheron. 2008. Complete mitochondrial genome sequences of the Arctic Ocean codfishes *Arctogadus glacialis* and *Boreogadus saida* reveal oriL and tRNA gene duplications. Polar Biology 31:1245–1252.

Bryant, D., and V. Moulton. 2002. NeighborNet: An agglomerative method for the construction of planar phylogenetic networks. In R. Guigo and D. Guseld, editors, *Algorithms in Bioinformatics*, WABI 2002, volume LNCS 2452, pages 375–391.

Charlesworth, B., R. Lande and M. Slatkin. 1982. A neo-Darwinian commentary on macroevolution. Evolution 36:474-498.

Chevreux, B., T. Wetter, and S. Suhai. 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) 99:45–56.

Comiso, J.C., C.L. Parkinson, R. Gersten, and L. Stock. 2008. Accelerated decline in the Arctic sea ice cover. Geophysical Research Letters 35: L01703.

Conant, G.C., G.P. Wagner, P.F. Stadler. 2007. Modeling amino acid substitution patterns in orthologous and paralogous genes. Molecular Phylogenetics and Evolution 42:298–307.

Correa, C.C., W.C. Aw, R.G. Melvin, N. Pichaud, and J.W.O. Ballard. 2012. Mitochondrial DNA variants influence mitochondrial bioenergetics in Drosophila melanogaster. Mitochondrion 12:459–464.

DaCosta, J.M. and M.D. Sorenson. 2014. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS ONE 9(9): e106713.

Delport, W., A.F. Poon, S.D. Frost, and S.L. Kosakovsky Pond. 2010. Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457.

Dowling, T.E., D.F. Markle, G.J. Tranah, E.W. Carson, D.W. Wagman, and B.P. May. 2016. Introgressive hybridization and the evolution of lake-adapted catostomid fishes. PLoS ONE 11: e0149884.

Dray, S. and A-B. Dufour. 2007. The ade4 package: Implementing the duality diagram for ecologists. Journal of Statistical Software 22:1–20.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 19:1792–1797.

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

Evanno, G., Regnaut, S., and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology 14:2611–2620.

Excoffier, L.P., E. Smouse, and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491.

Excoffier, L., and H.E.L. Lischer. 2010. Arlequin Suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564–567.

Fevolden, S., I. Martinez, and J. Christiansen. 1999. RAPD and scnDNA analyses of polar cod, *Boreogadus saida* (Pisces, Galidae), in the north Atlantic. Sarsia 84:99–103

Foote, A.D., P.A. Morin, J.W. Durban, R.L. Pitman, P. Wade, E. Willerslev, M.T.P. Gilbert, and R.R.D. Fonseca. 2010. Positive selection on the killer whale mitogenome. Biology Letters 7:116–118.

Fonseca, R.R.D., W.E. Johnson, S.J. O'Brien, M.J. Ramos, and A. Antunes. 2008. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics 9:119.

Fossheim, M., R. Primicerio, E. Johannesen, R.B. Ingvaldsen, M.M. Aschan, and A.V. Dolgov. 2015. Recent warming leads to a rapid borealization of fish communities in the Arctic. Nature Climate Change 5: 673–677.

Frainer, A., R. Primicerio, S. Kortsch, M. Aune, A.V. Dolgov, M. Fossheim, and M.M. Aschan. 2017. Climate-driven changes in functional biogeography of Arctic fish communities. Proceedings of the National Academy of Sciences of the United States: doi: 10.1073/pnas.1706080114.

Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selections. Genetics 147:915–925.

Galtier, N., B. Nabholz, S. Glemin, and G. Hurst. 2009. Mitochondrial DNA as a marker of molecular diversity: A reappraisal. Molecular Ecology 18:4541–4550.

Garvin, M.R., J.P. Bielwaski, and A.J. Gharrett. 2011. Positive Darwinian selection in the piston that powers proton pumps in Complex I of the mitochondria of Pacific Salmon. PLoS ONE 6: e24127.

Gittleman, R. M., J. G. Schraiber, B. Vernot, C. Mikacenic, M. M. Wurfel, and J. M. Akey. 2016. Archaic hominin admixture facilitated adaptation to out-of-Africa environments. Current Biology 26:3375–3382. doi:10.1016.jcub.2016.10.041.

Goudet, J., M. Raymond, T. De Meeus, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics 144:1933–1940.

Hahn, C., L. Bachmann, and B. Chevreux. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative approach. Nucleic Acids Research 41: e129.

Harrison, K. A. Pavlova, H.M. Gan, Y.P. Lee, C.M. Austin, and P. Sunnucks. 2016. Pleistocene divergence across a mountain range and the influence of selection on mitogenome evolution in threatened Australian freshwater cod species. Heredity 116:506–515.

Hedrick, P. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology 22:4606–4618.

Hill, G.E. 2015. Mitonuclear ecology. Molecular Biology and Evolution 32:1917–1927.

Holldórsdóttir, K. and E. Árnason 2015. Whole-genome sequencing uncovers cryptic and hybrid species among Atlantic and Pacific cod-fish. bioRxiv: doi:https://doi.org/10.1101/034926.

Howell, S.E.L., M. Brady, C. Derksen, and R.E. Kelly. 2016. Recent changes in sea ice area flux through the Beaufort Sea during the summer. Journal of Geophysical Research 121:2659–2672.

Hubisz, M.J., D. Falush, M. Stephens, and J.K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322–1332.

Huson, D.H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23:254–267. Software available from www.splitstree.org.

Iwasaki, W., T. Fukunaga, R. Isagozawawa, K. Yamada, Y. Maeda, T.P. Satoh, T. Sado, K. Mabuchi, H. Takeshima, M. Miya, and M Nishida. 2013. MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. Molecular Biology and Evolution 30:2531–2540.

Jakobsdóttir, K.B., Þ. D. Jörundsdóttir, S. Skírnisdóttir, S. Hjörleifsdóttir, G.Ó. Hreggviðsson, A.K. Daníelsdóttir, and C. Pampoulie. 2006. Nine new polymorphic microsatellite loci for the amplification of archived otolith DNA of Atlantic cod, *Gadus morhua* L. Molecular Ecology Notes 6:337–339.

Jombart, T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405.

Jørgensen, L.L., P. Archambault, C. Armstrong, A. Dolgov, E. Edinger, T. Gaston, J. Hildebrand, D. Piepenburg, W. Smith, C. von Quillfeldt, M. Vecchione, and J. Rice. 2016. Arctic Marine biodiversity. In First Global Marine Assessment, Chapter 36G. *Oceans and Law of the Sea*, United Nations. Pp.47.

Karlsen B.O., A. Emblem, T.E. Jørgensen, K.A. Klingan, J. T. Nordeide, T. Moum, and S.D. Johansen. 2014. Mitogenome sequence variation in migratory and stationary ecotypes of North-east Atlantic cod. Marine Genomics 15:103–108.

Kingsolver, J. G., H. E. Hoekstra, J. Hoestra, D. Berrigan, S. N. Vignieri, C.E. Hill, A. Hoang, P. Gilbert and P. Beerli. 2001. The strength of super genetic selection in natural populations. The American Naturalist 157:245-261.

Kosakovsky Pond, S.L., S.D.W. Frost, and S.V. Muse. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679.

Lane, N., and W. Martin. 2010. The energetics of genome complexity. Nature 467:929–934.

Lanfear, R., P.B. Frandsen, A.M. Wright, T. Senfeld, and B. Calcott. 2016. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34:772–773.

Laurel, B.J., M. Spencer, P. Iseri, and L.A. Copeman. 2016. Temperature-dependent growth and behavior of juvenile Arctic cod (*Boreogadus saida*) and co-occurring North Pacific gadids. Polar Biology 39:1127–1135

Lavretsky, P., J.L. Peters, K. Winker, V. Bahn, I. Kulikova, Y.N. Zhuravlev, R.E. Wilson, C. Barger, K. Gurney, and K.G. McCracken. 2016. Becoming pure: identifying generational classes of admixed individuals within lesser and greater scaup populations. Molecular Ecology 25:661–674.

Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Machado, D.J., M.L. Lyra, and T. Grant. 2016. Mitogenome assembly from genomic multiplex libraries: Comparison of strategies and novel mitogenomes for five species of frogs. Molecular Ecology Resources 16:686–693.

Madsen, M.L., S-E. Fevolden, and J.S. Christiansen. 2009. A simple molecular approach to distinguish between two Arctic gadoid fishes *Arctogadus glacialis* (Peters, 1874) *Boreogadus saida* (Lepechin, 1774). Polar Biology 32:937–939.

Madsen, M.L., R.J. Nelson, S-E. Fevolden, J.S. Christiansen, and K. Praebel. 2016. Population genetic analysis of Euro-Arctic polar cod Boreogadus saida suggests fjord and oceanic structuring. Polar Biology 39:969–980.

MBC Applied Environmental Sciences. 2003. Physical Oceanography of the Beaufort Sea – Workshop Proceedings. OCS Study MMS 2003-045. Prepared by MBC Applied Environmental Sciences, Costa Mesa, CA. Prepared for the U.S. Dept. of the Interior, Minerals Management Service, Alaska OCS Region, Anchorage, AK. 26 pp. plus attachments.

McClellan, D.A., E.J. Palfreyman, M.J. Smith, J.L. Moss, R.G. Christensen, and J.K. Sailsbery. 2005. Physiochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. Molecular Biology and Evolution 22: 437–455.

McDonald, J.H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351:652–654.

Medrano J. F., E. Aasen, and J. Sharrow. 1990. DNA extraction from nucleated red blood cells. Biotechniques 8:43.

Miller, K.M., K.D. Le, and T.D. Beacham. 2000. Development of tri- and tetranucleotide repeat microsatellite loci in Atlantic cod (*Gadus morhua*). Molecular Ecology 9:237–244.

Mjelle, K.A., B.O. Karlsen, T.E. Jørgensen, T. Moum, and S.D. Johansen. 2008. Halibut mitochondrial genomes contain extensive heteroplasmic tandem repeat arrays involved in DNA recombination. BMC Genomics 9:10.

Morgan-Richards, M., M. Bulgarella, L. Sivyer, E.J. Dowle, M. Hale, N.E. McKean, and S.A. Trewick. 2017. Explaining large mitochondrial sequence differences within a population sample. Royal Society Open Science 4(11):170730.

Muhlfeld, C. C., S.T. Kalinowski, T.E. McMahon, M.L. Taper, S. Painter, R.F. Leary, and F. Allendorf, F. W. 2009. Hybridization rapidly reduces fitness of a native trout in the wild. Biology Letters 5:328–331.

Murrell, B.J.O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S.L. Kosakovsky Pond. 2012. Detecting individual sites subject to episodic diversifying selection. PloS Genetics 8:e1002764.

Murrell, B., S. Moola, A. Mabona, T. Weighill, D. Sheward, S.L. Kosakovsky Pond, and K. Scheffler. 2013. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection Molecular Biology and Evolution 30:1196–1205.

Nabholz, B., H. Ellegren, and J.B Wolf. 2013. High levels of gene expression explain the strong evolutionary constraint of mitochondrial protein-coding genes. Molecular Biology and Evolution 30:272–284.

Nelson, R.J., C. Bouchard, M. Madsen, K. Praebel, E. Rondeau, K. von Schalburg, J.S. Leong, S. Jantzen, Z. Sandwith, S. Puckett, A. Messmer, S-E. Fevolden, and B.F. Koop. 2013. Microsatellite loci for genetic analysis of the Arctic gadids *Boreogadus saida* and *Arctogadus glacialis*. Conservation Genetic Resources 5:445-448.

O'Reilly P.T., M.F. Canino, K.M. Bailey, and P. Bentzen. 2000. Isolation of twenty low stutter di- and tetranucleotide microsatellites for population analyses of walleye pollock and other gadoids. Journal of Fish Biology 56:1074–1086.

Pálsson, S., T. Källman, J. Paulsen, and E. Árnason. 2009. An assessment of mtDNA variation in Arctic gadoids. Polar Biology 32:471–479.

Pampoulie C, S. Skirnisdottir, B. Star, S. Jentoft, I. G. Jónsdóttir, E. Hjörleifsson, V. Thorsteinsson, Ó.K. Pálsson, P.R. Berg, Ø. Andersen, S. Magnusdottir, S.J. Helyar, and A.K. Daníelsdóttir. 2015. Rhodopsin gene polymorphism associated with divergent light environments in Atlantic cod. Behavior Genetics 45:236–244.

Pavlova, A., H.M. Gan, Y.P. Lee, C.M. Austin, D.M. Gilligan, M. Lintermans, and P. Sunnucks. 2017. Purifying selection and genetic drift shaped Pleistocene evolution of the mitochondrial genome in an endangered Australian freshwater fish. Heredity 118:466-476.

Petit, R.J., A. El Mousadik, and O. Pons. 1998. Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12:844–855.

Pfeifer, B., U. Wittelsbürger, S.E. Ramos-Onsins, and M.J. Lercher. 2014. PopGenome: an efficient Swiss army knife for population genomics analyses in R. Molecular Biology and Evolution 31:1929–1936.

Pilgrim, B.L., R.C. Perry, J.L. Barron, and H.D. Marshall. 2008. Nucleotide variation in the mitochondrial genome provides evidence for dual routes of postglacial recolonization and genetic recombination in the northeastern brook trout (*Salvelinus fontinalis*). Genetics and Molecular Research 11:3466–3481.

Pritchard, J., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Prost, S. and C.N.K. Anderson. 2011. TempNet: a method to display statistical parsimony networks for heterochronous DNA sequence data. Methods in Ecology and Evolution 2:663e667.

Ralph, P., and G. Coop. 2010. Parallel adaptation: one or many waves of advance of an advantageous allele? Genetics 186: 647–668

Rambaut, A. 2009. FigTree version 1.3.1 [computer program]. Available from http://tree.bio.ed.ac.uk.

Rambaut, A., M.A. Suchard, D. Xie, and A.J. Drummond. 2014. Tracer v1.6 [computer program]. Available from http://tree.bio.ed.ac.uk/software/tracer.

Rand, D.M., R.A. Haney, and A.J. Fry. 2004. Cytonuclear coevolution: The genomics of cooperation. Trends in Ecology and Evolution 19:646–653.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248–249.

Renaud, P.E., J. Berge, Varpe, O.J. Lonne, J. Nahrgang, C. Ottesen, and I. Hallanger. 2012. Is the poleward expansion by Atlantic cod and haddock threatening native polar cod, *Boreogadus saida*. Polar Biology 35: 401–412.

Roa-Varón, A., and G. Ortí. 2009. Phylogenetic relationships among families of Gadiformes (Teleostei, Paracanthopterygii) based on nuclear and mitochondrial data. Molecular Phylogenetics and Evolution 52:688–704.

Ronquist, F., M. Teslenko, P. van der Mark, D.L. Ayres, A. Darling, S. Höhna, B. Larget, L. Lui, M.A. Suchard, and J.P. Huelsenback. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–542.

Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources 8:103–106.

Salmenkova, E.A. 2011. New view on the population genetic structure of marine fish. Russian Journal of Genetics 47:1279–1287.

Sá-Pinto, A, M.S. Branco, P.B. Alexandrino, M.C. Fontaine, and S.J.E. Baird. 2012. Barriers to gene flow in the marine environments: Insights from two common intertidal limpet species of the Atlantic and Mediterranean. PLoS ONE 7:e50330.

Scott, G. R., P.M. Schulte, S. Egginton, J.G. Richards, and W.K. Milsom. 2010. Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. Molecular Biology and Evolution 28:351–363.

Sefc, K. M., R. B. Payne, and M. D. Sorenson. 2003. Microdatellite amplification from museum feather samples: Effects of fragment size and template concentration on genotyping errors. The Auk 120:982-989.

Shum, P., C. Pampoulie, C. Sacchi, and S. Mariani. 2014. Divergence by depth in an oceanic fish. PeerJ 2:e525.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.

Skirnisdottir, S., C. Pampoulie, S. Hauksdottir, I. Schulte, K. Olafsson, G.O. Hreggvidsson, and S. Hjorleifdottir. 2008. Characterization of 18 new microsatellite loci in Atlantic cod (*Gadus morhua* L.). Molecular Ecology Resources 8:1503–1505.

Smoot, C.A. and R.R. Hopcroft. 2017. Cross-shelf gradients of epipelagic zooplankton communities of the Beaufort Sea and the influence of localized hydrographic features. Journal of Plankton Research 39:65–78.

Sonsthagen, S.A., S.L. Talbot, and C.M. White. 2004. Gene flow and genetic characterization of Northern Goshawks breeding in Utah. Condor 106:826–836.

Stabeno, P.J., N.B. Kachel, S.E. Moore, J.M. Napp, M. Sigler, A. Yamaguchi, and A.N. Zerbini. 2012. Comparison of warm and cold years on the southern Bering Sea shelf and some implications for the ecosystem. Deep Sea Research Part II: Topical Studies in Oceanography 65-70:31–45.

Talbot, S.L., A. Palmer, G.K. Sage, S. Sonsthagen, T. Swem, D. Brimm, and C.M. White. 2011. Lack of genetic polymorphism among peregrine falcons of Fiji. Journal of Avian Biology 42:415–428

Tajima, F. 1989. The effect of change in population size on DNA polymorphism. Genetics 123:597–601.

Templeton, A.R. 1989. The meaning of species and speciation: a genetic perspective. In: D. Otte and J.A. Endler, editors. *Speciation and its consequences*. Sunderland: Sinauer Associates; pp. 3–27.

Thorsteinson, L.K., and Love, M.S., eds. 2016. Alaska Arctic marine fish ecology catalog: U.S. Geological Survey Scientific Investigations Report 2016-5038 (OCS Study, BOEM 2016-048), 768 p., http://dx.doi.org/10.3133/sir20165038.

Van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills, and P. Shipley. 2004. MICROCHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535–538.

Wang, M., J.E. Overland, and P. Stabeno. 2012. Future climate of the Bering and Chukchi Seas projected by global climate models. Deep Sea Research Part II: Topical Studies in Oceanography 65-70:46–57.

Wataru, I., T. Fukunaga, R. Isagozawa, K. Yamada, Y. Maeda, T.P. Satoh, T. Sado, K. Mabuchi, H. Takeshima, M. Miya, and M. Nishida. 2013. MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. Molecular Biology and Evolution 30:2531–2540.

Weir, B.S. and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

White, D.J., J.N. Wolff, M. Pierson, and N.J. Gemmell. 2008. Revealing the hidden complexities of mtDNA inheritance. Molecular Ecology 17:4925–4942.

Wildes, S.L., J. Whittle, H. Nguyen, and J. Guyon. 2016. *Boreogadus saida* genetics in the Alaskan Arctic. US Dept. of the Interior, Bureau of Ocean Energy Management, Alaska OCS Region. OCS Study BOEM 2011-AK-11-08 a/b. 67 pp. – DRAFT REPORT.

Wilson, R.E., B.J. Pierson, G.K. Sage, M.C. Gravley, D.M. Menning. S.A. Sonsthagen, J. Nelson, and S.L. Talbot. 2018. Genetic and genomic data on Arctic cod (*Boreogadus saida*) and other gadids, Beaufort Sea, Alaska, 2011-2014. U.S. Geological Survey data release. http://doi.org/10.5066/F7CF9P23.

Woolley, S., J. Johnson, M.J. Smith, K.A. Crandall, and D.A. McClellan. 2003. TreeSAAP: Selection on amino acid properties using phylogenetic trees. Bioinformatics 19:671–672.

Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. Computer Applications in Biosciences 13:555–556.

Zhang, J. and R.E. Broughton. 2013. Mitochondrial-nuclear interactions: compensatory evolution or variable functional constraints among vertebrate oxidative phosphorylation genes? Genome Biology and Evolution 5: 1781–1791.

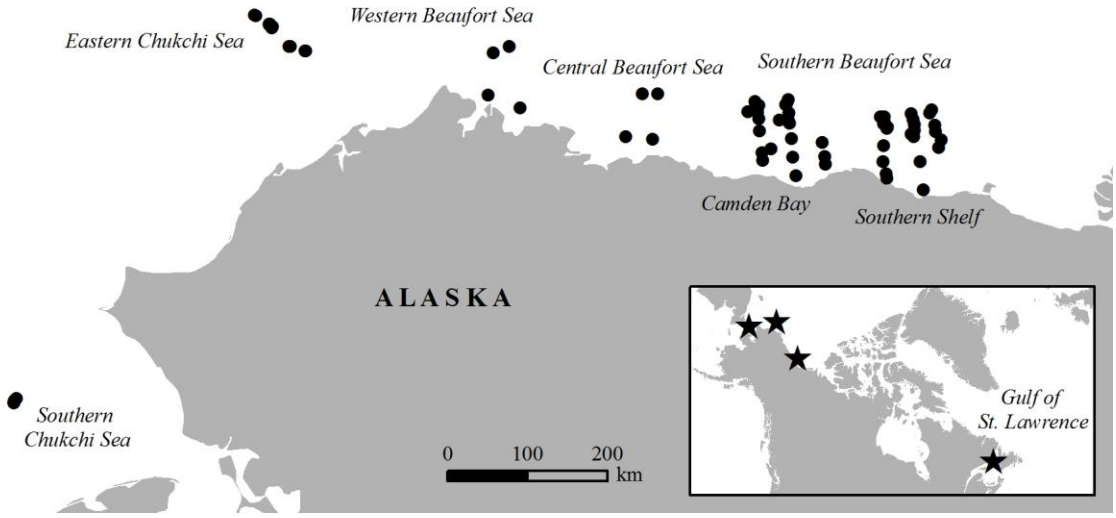


Figure 2-1. Map of Alaska and Canada, with detail of general sampling locales of Arctic cod indicated by stars in the insert, and Beaufort and Chukchi sea sampling locales represented by black dots.

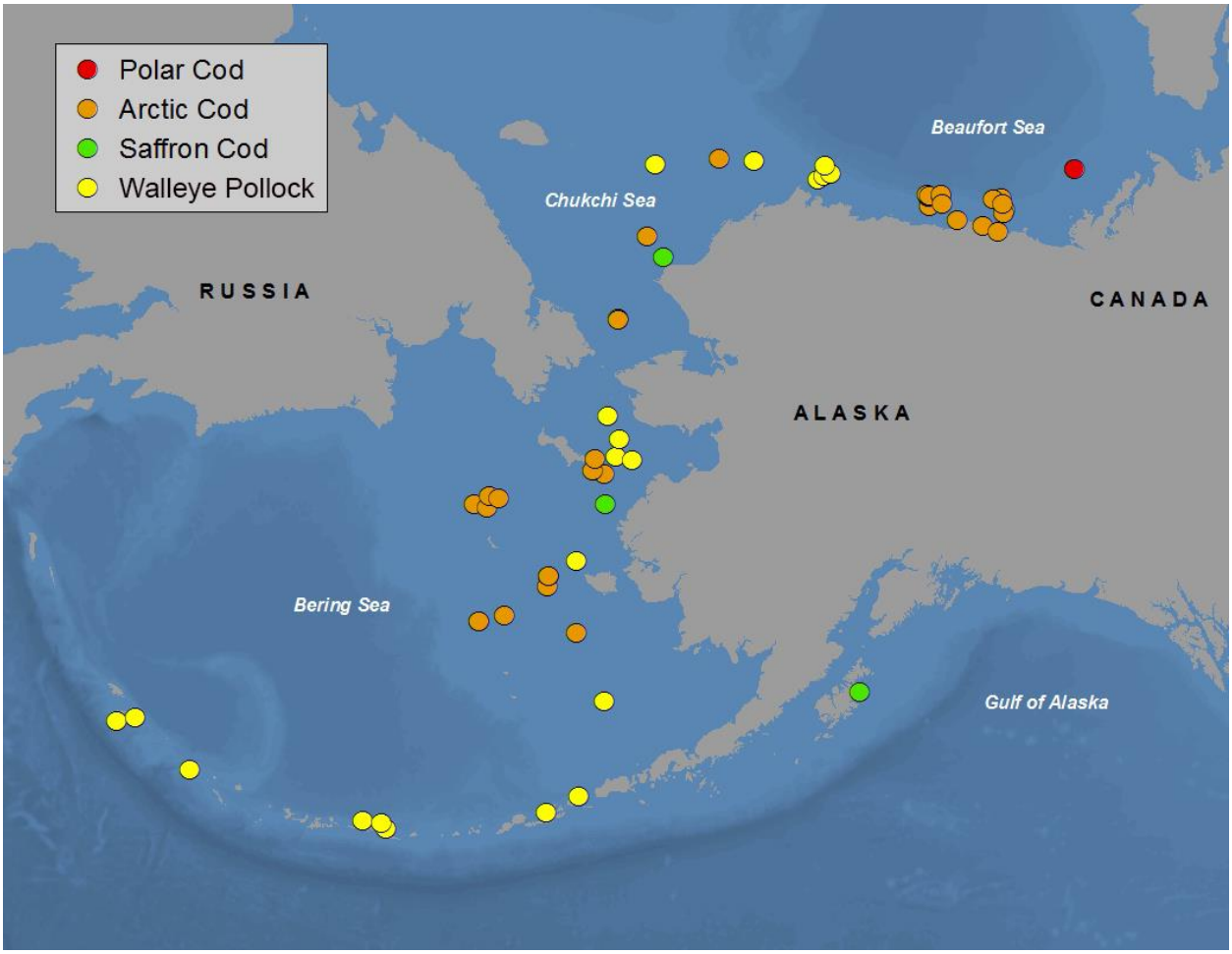


Figure 2-2. Map of the Bering, Chukchi and Beaufort Sea region showing general sampling locales for each of the four gadid species (indicated by different colored dots).

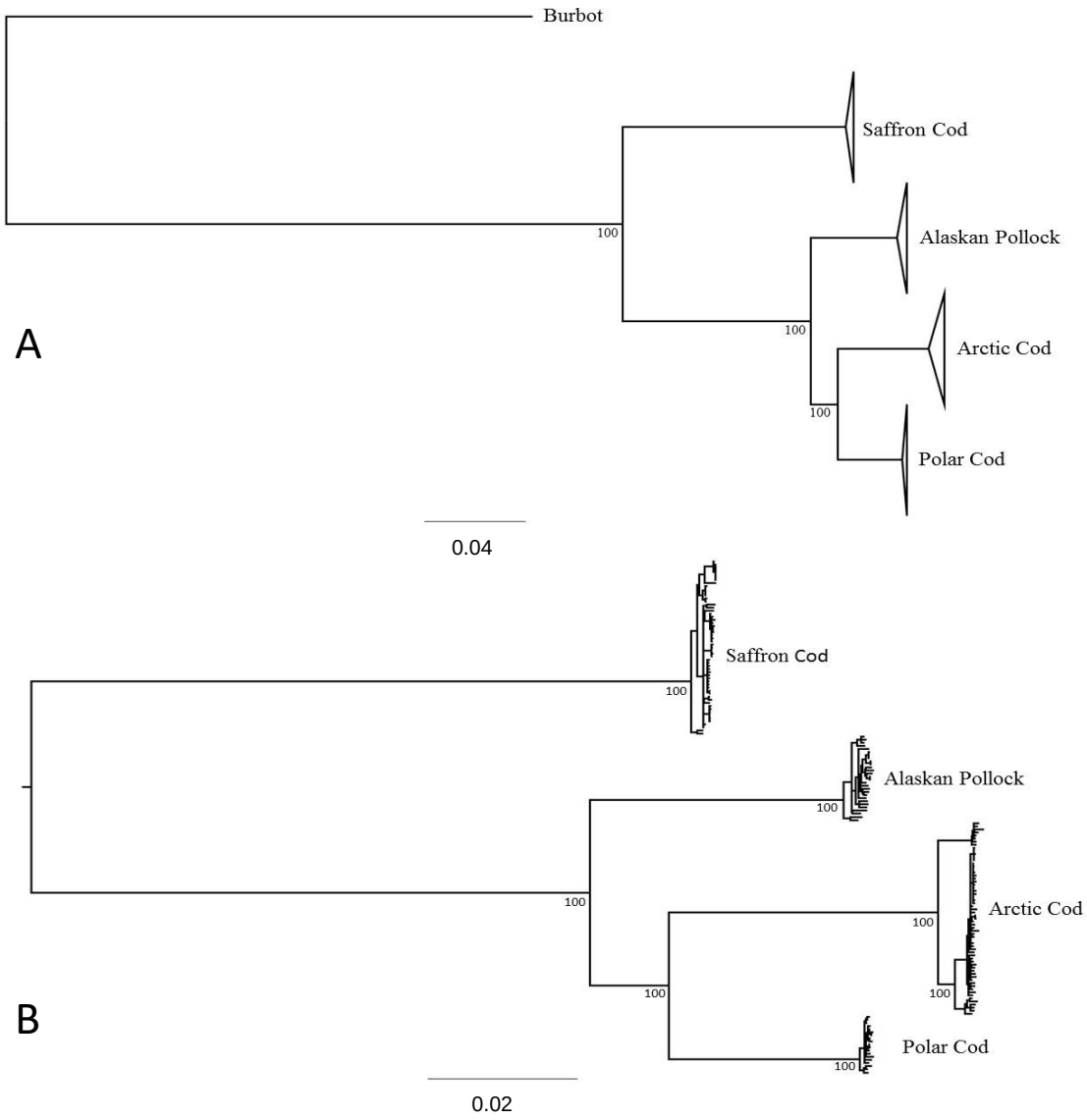


Figure 2-3. Phylogenetic trees showing relationships among the four gadid species (A) with an outgroup (Burbot, *Lota lota*; GenBank accession NC004379) and (B) with outgroup removed to illustrate partitions within Arctic cod and to conduct codon-based selection analyses. Phylogenetic trees are derived using Bayesian phylogenetic analysis (Ronquist et al. 2012) based on partial or entire mitochondrial DNA genomes the four gadid species. Shown are 50% majority rule consensus trees.

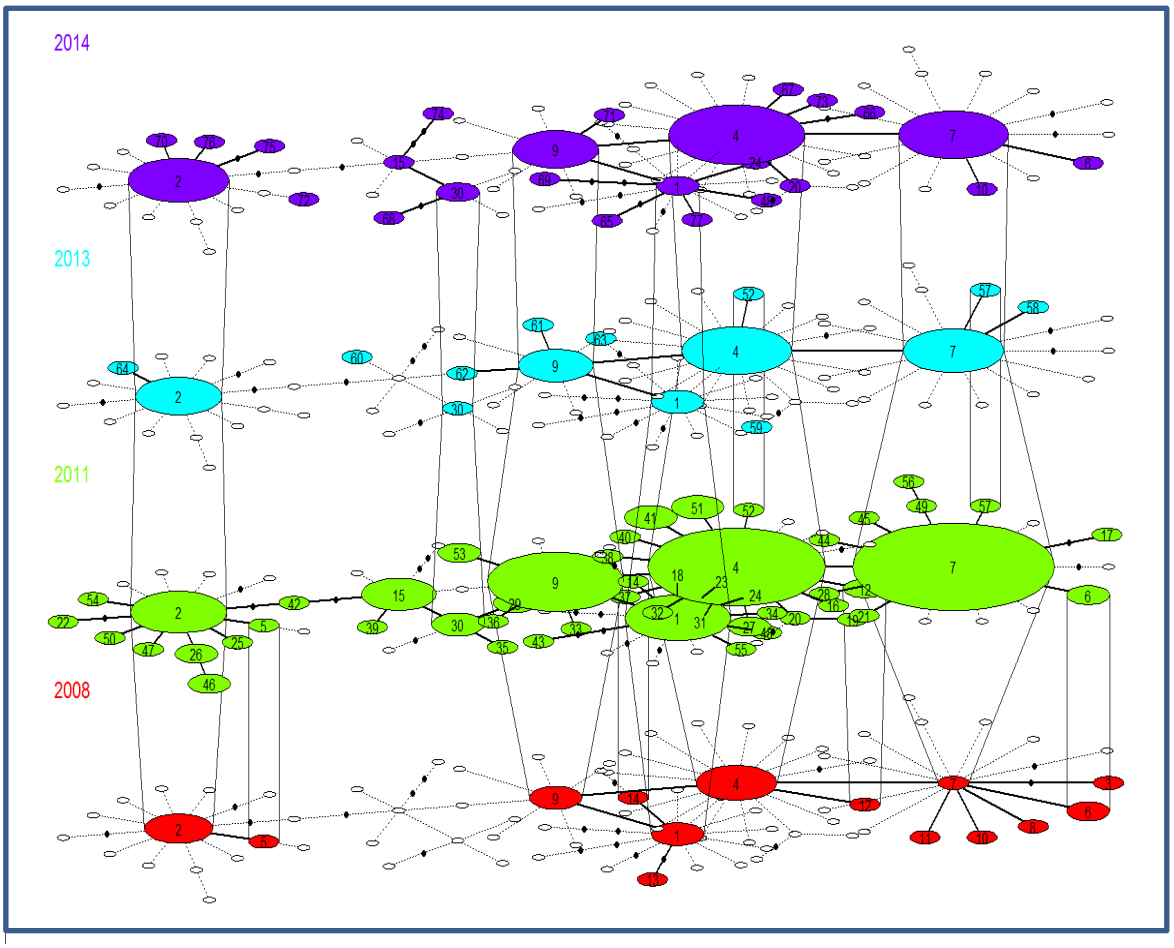


Figure 2-4. Temporal distribution of mitochondrial DNA cytochrome b haplotypes for Beaufort Sea samples of Arctic cod using TempNet (Prost and Anderson 2011).

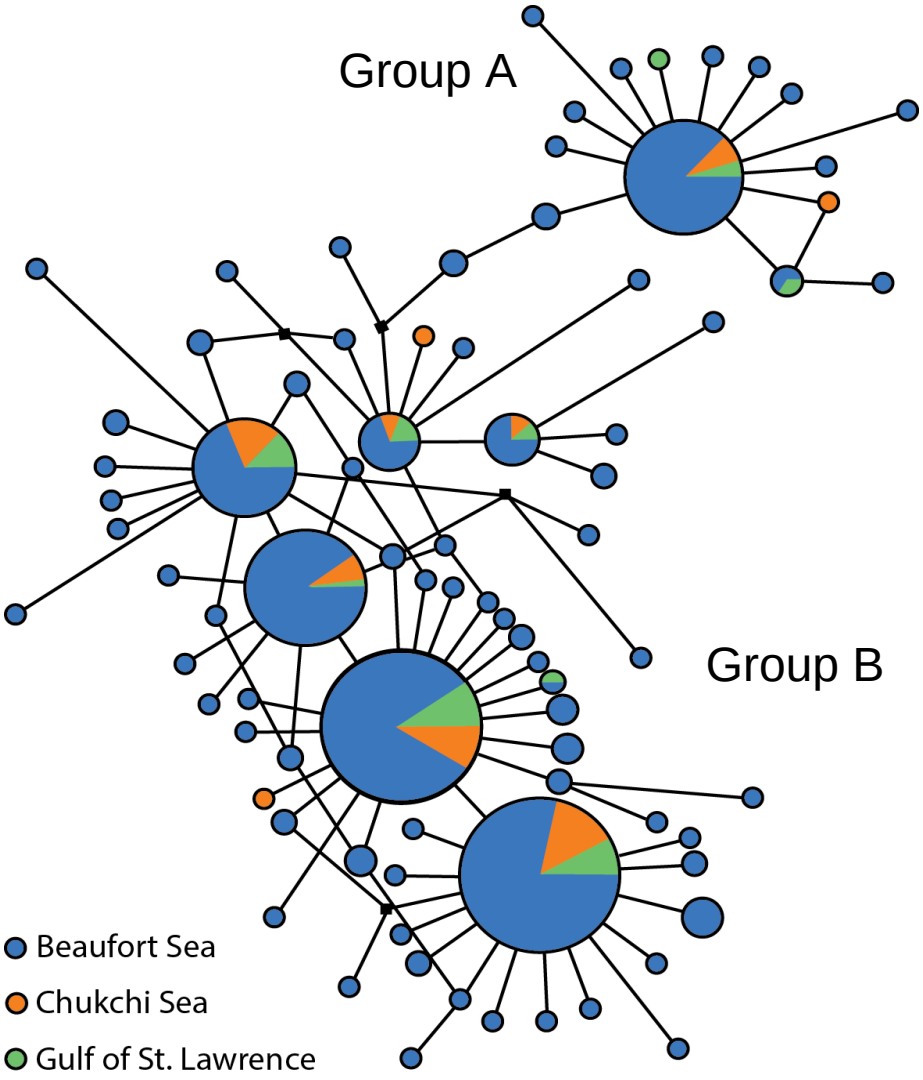


Figure 2-5. Parsimony network illustrating the relationships among 82 mitochondrial DNA cytochrome b haplotypes assayed from Arctic cod. Size of the node corresponds to the frequency of each haplotype. Beaufort Sea haplotypes are illustrated in blue, Chukchi Sea in orange, and Gulf of St. Lawrence in green. All line segments represent a difference at one nucleotide position between neighboring haplotypes. Black squares represent unsampled haplotypes.

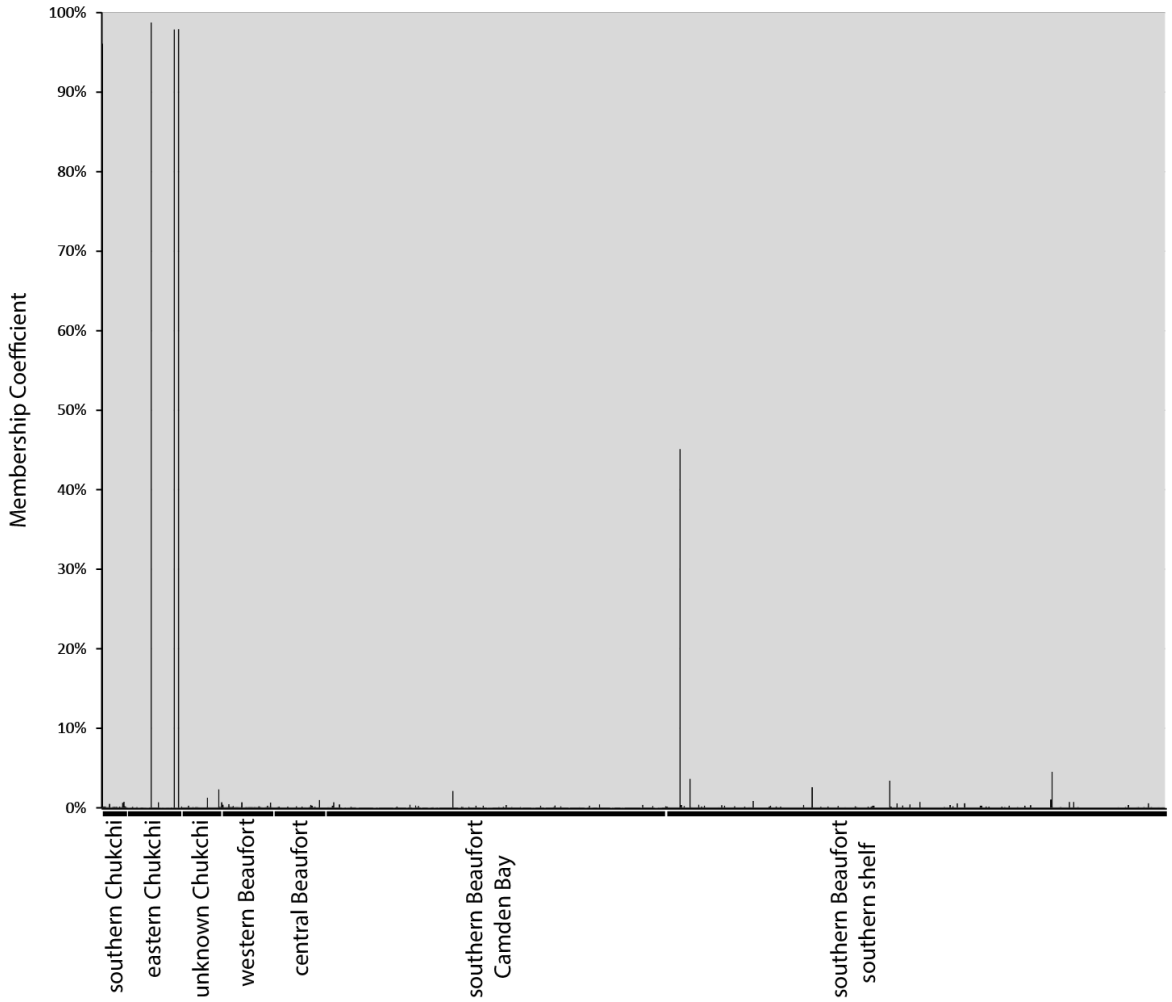


Figure 2-6. Results of Bayesian clustering analyses of microsatellite data. Individual samples are represented by a single vertical line along the x-axis, according to sampling region. The y-axis gives the probability of an individual assigning to each of the clusters (sum = 1). Two clusters (one represented by grey vertical lines, the other by black vertical lines) are supported (LnPR(K=2) = -20078.1, ΔK = 56.9).

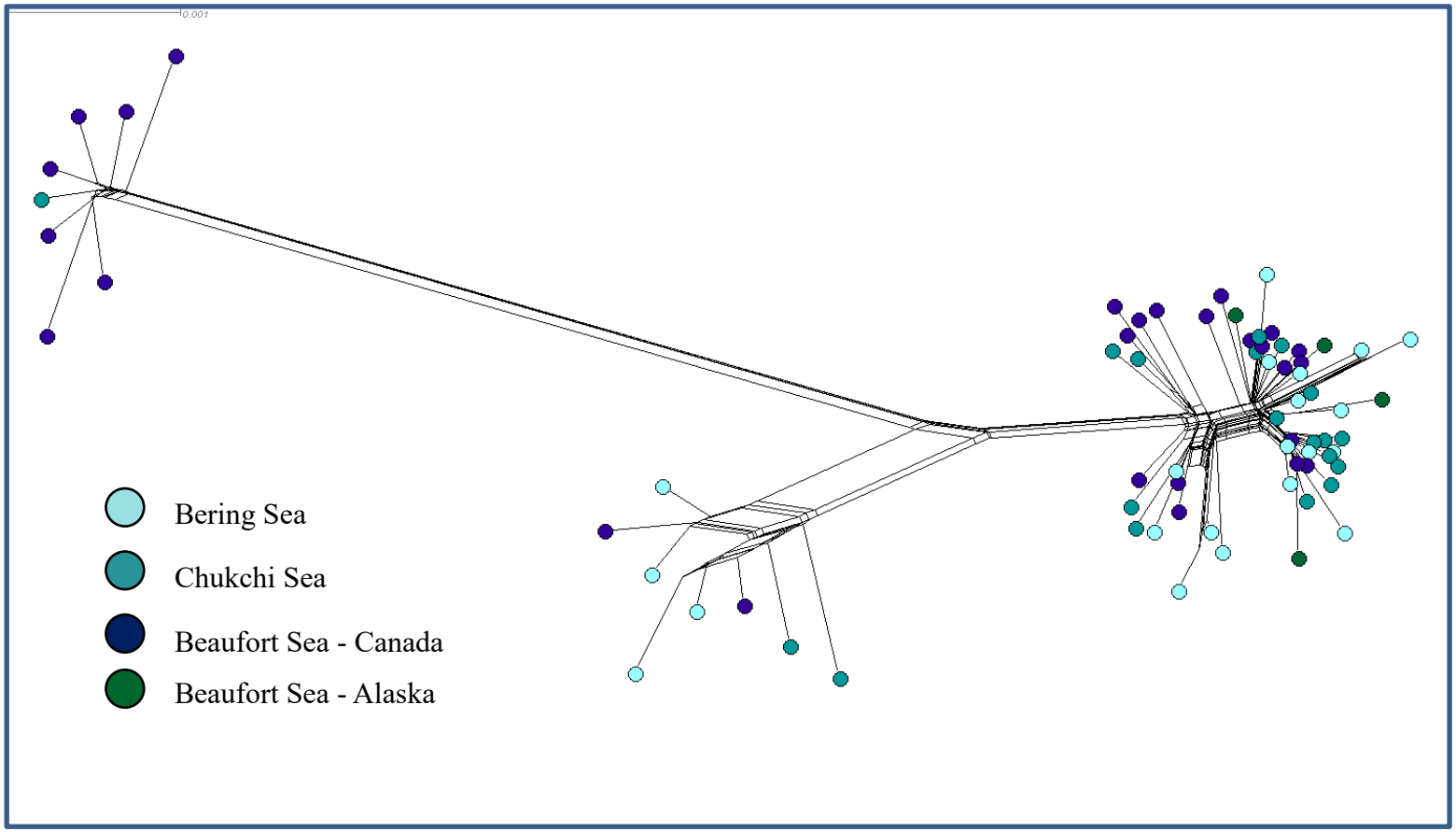


Figure 2-7. Neighbor-net tree showing relationships among mitogenome haplotypes within Arctic cod sampled from across the Beaufort, Chukchi and Bering seas.

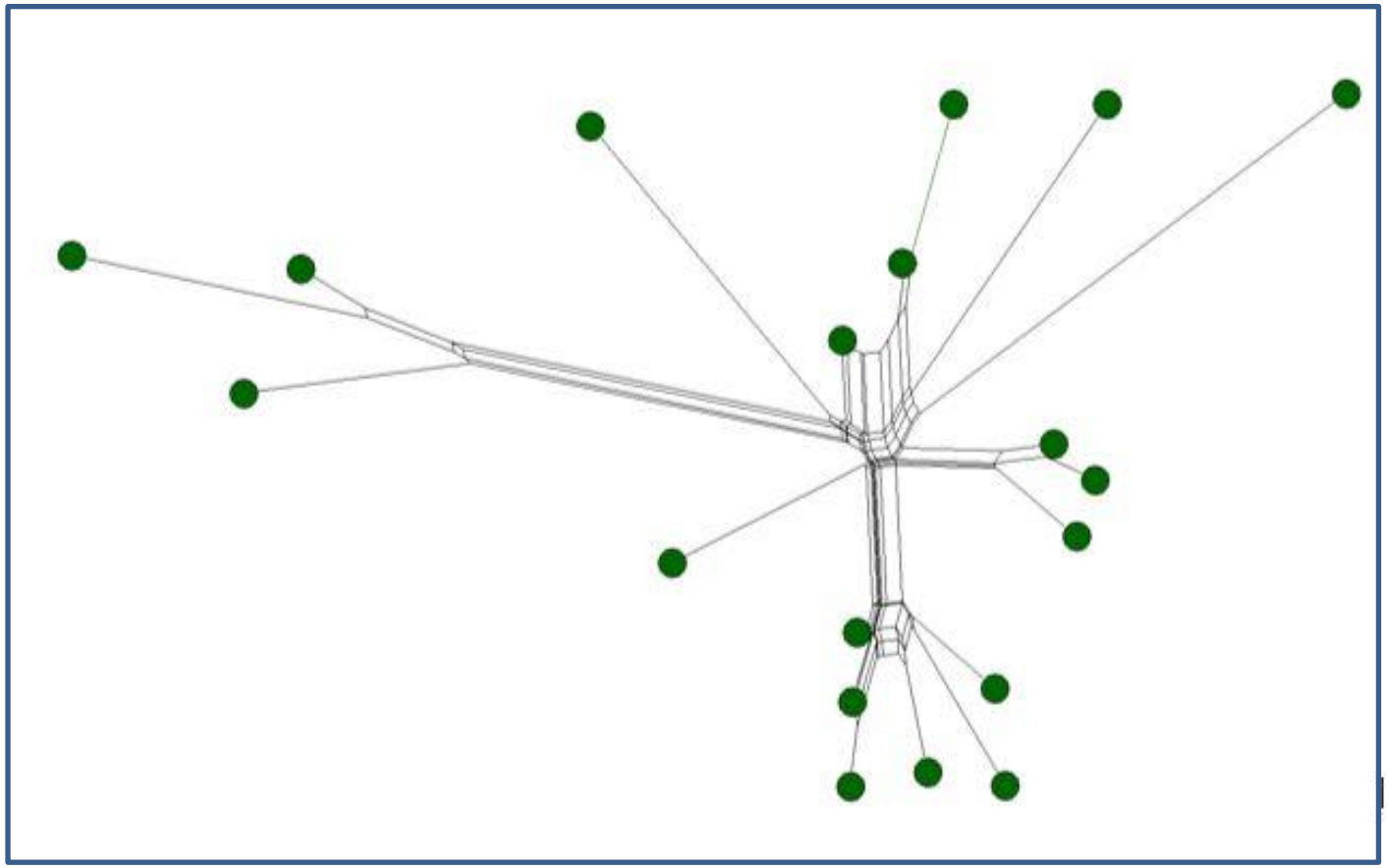


Figure 2-8. Neighbor-net tree showing relationships among mitogenome haplotypes within Polar cod sampled from the Canadian Beaufort Sea.

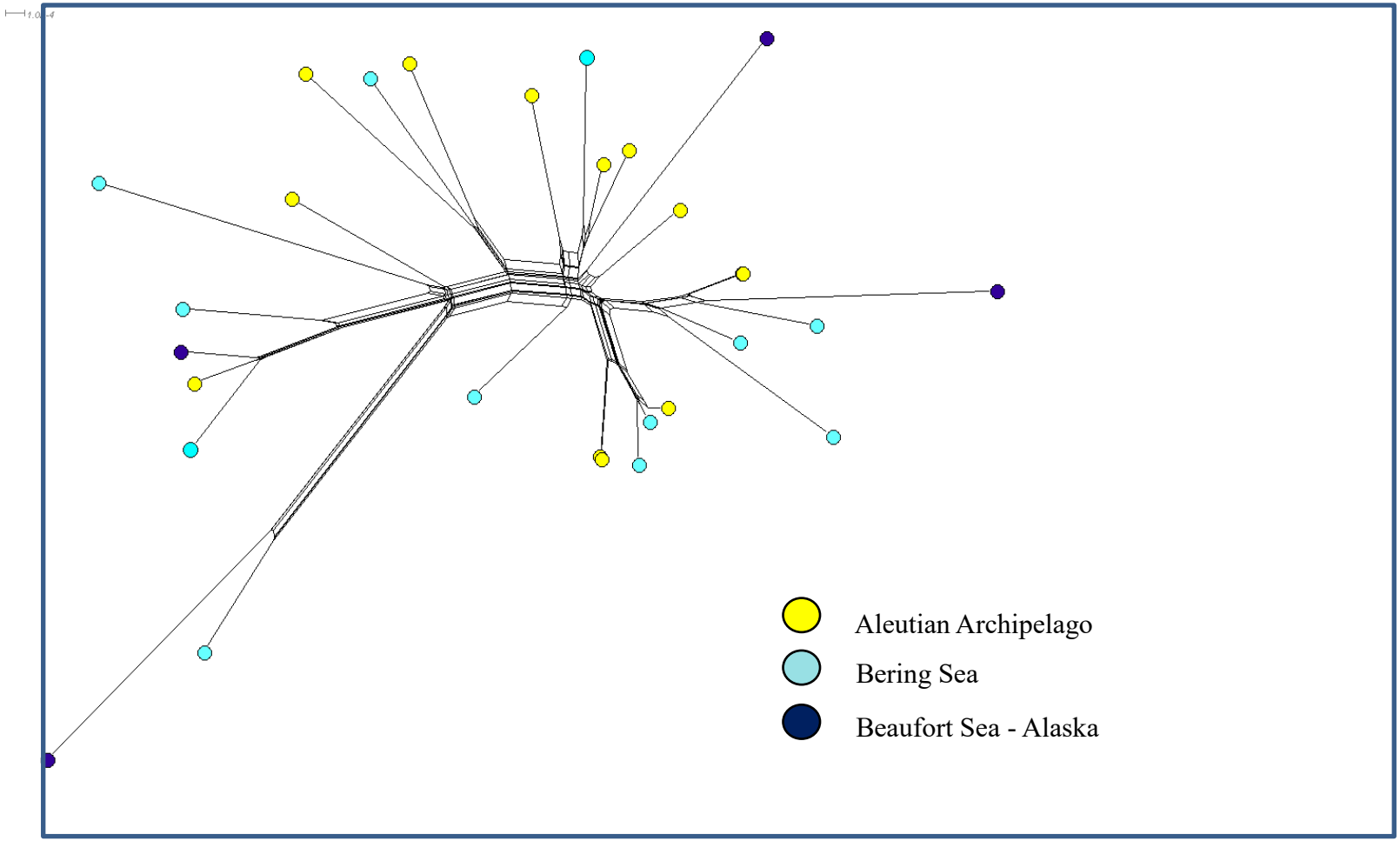


Figure 2-9. Neighbor-net tree showing relationships among mitogenome haplotypes within Walleye pollock sampled from the Bearing and Beaufort seas and the Aleutian Archipelago.

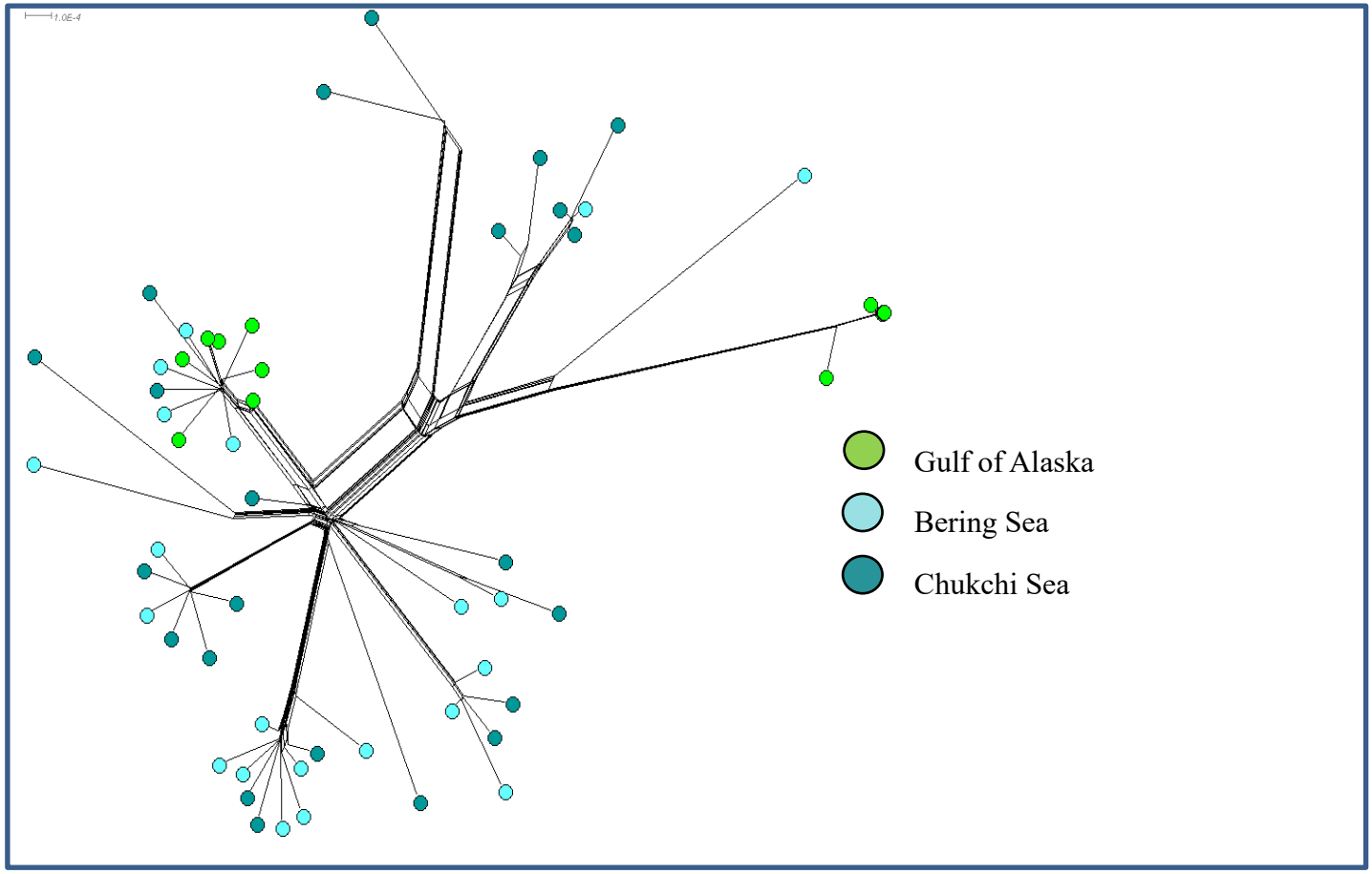


Figure 2-10. Neighbor-net tree showing relationships among mitogenome haplotypes within Saffron cod sampled from Bering and Chukchi seas and the Gulf of Alaska.

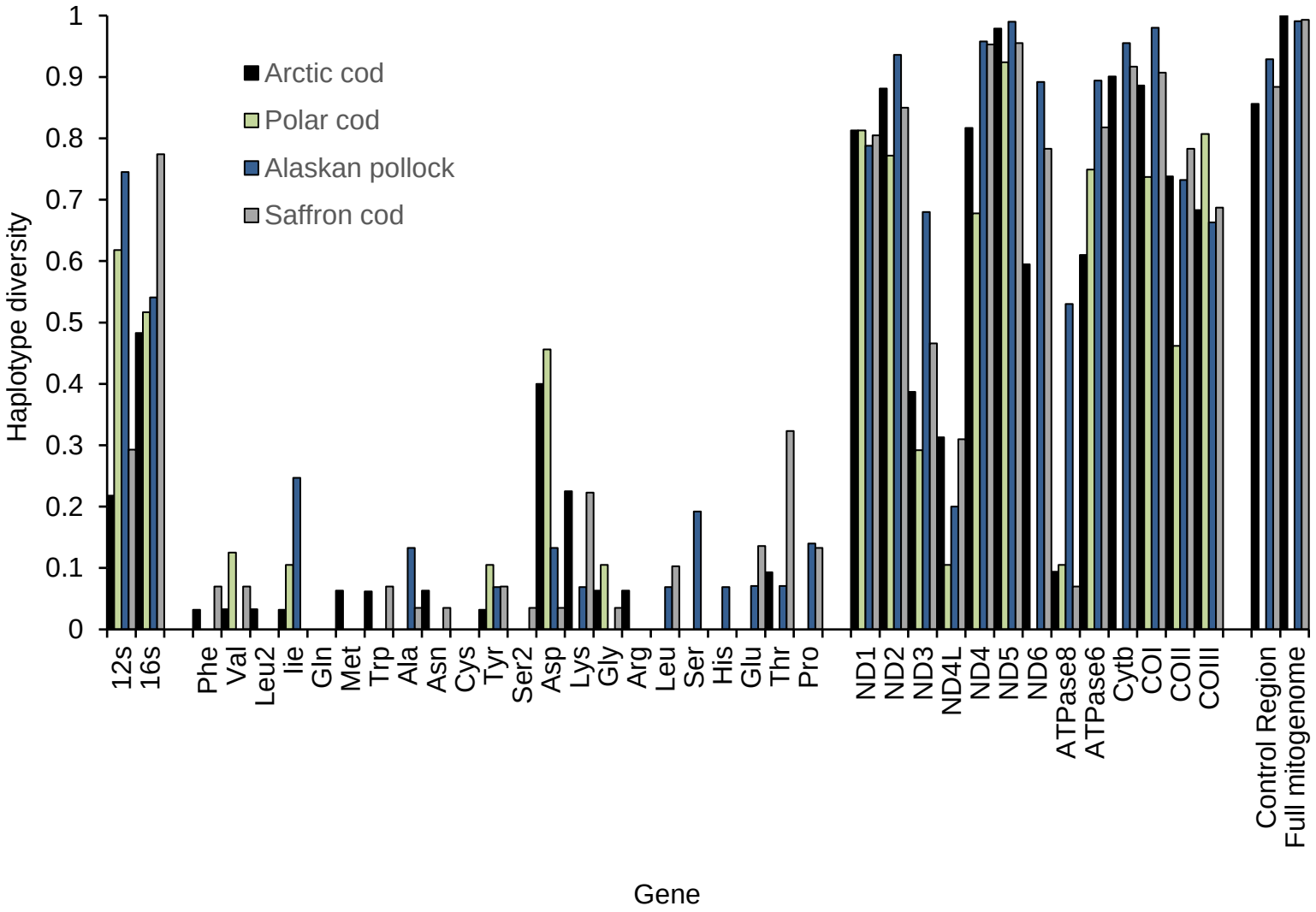


Figure 2-11. Haplotype diversity across mitochondrial genes for four gadid species sampled from Alaska and Canada.

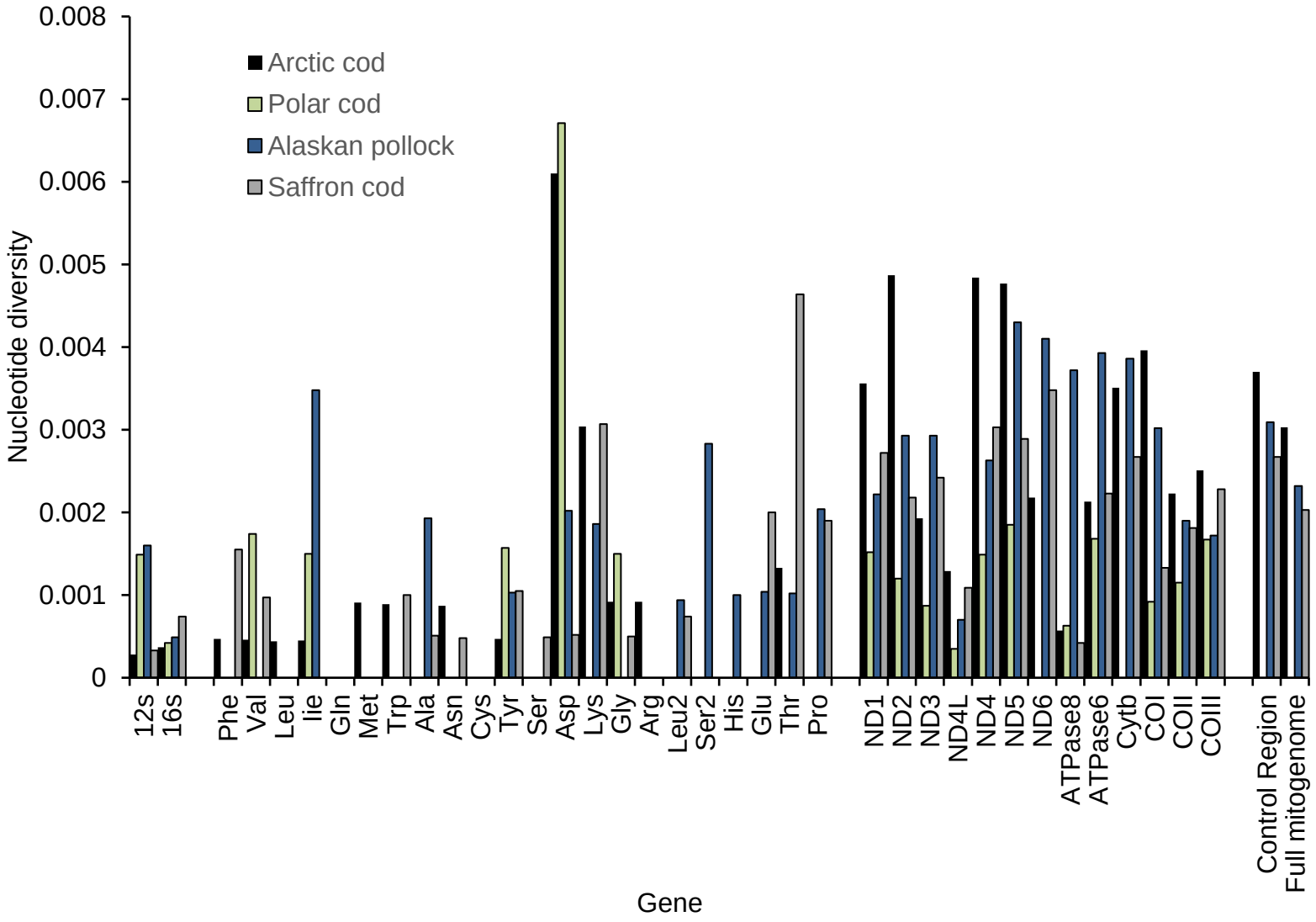


Figure 2-12. Nucleotide diversity across mitochondrial genes for four gadid species sampled from Alaska and Canada.

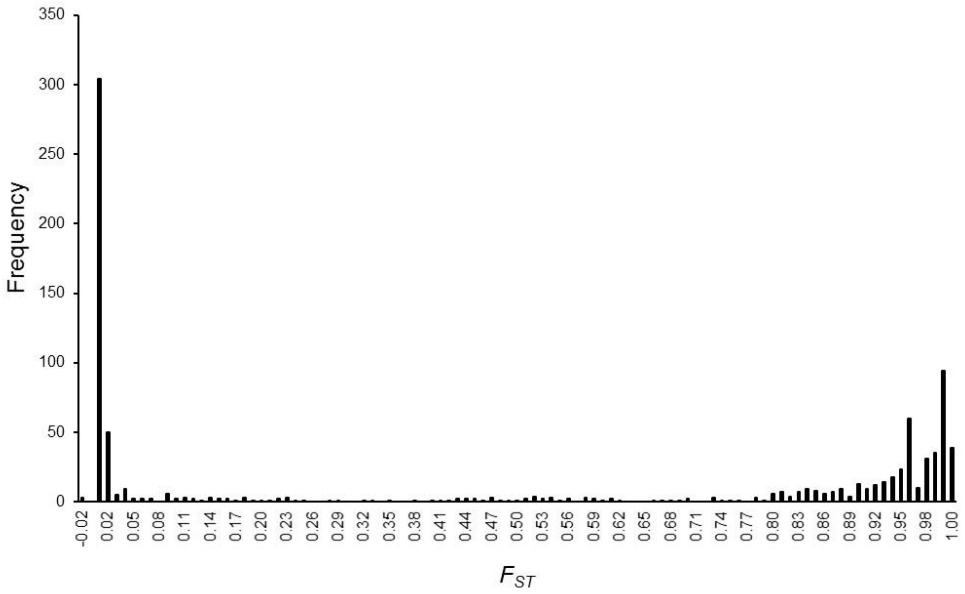


Figure 2-13. Frequency of F_{ST} estimates for 897 loci between Arctic cod and Polar cod.

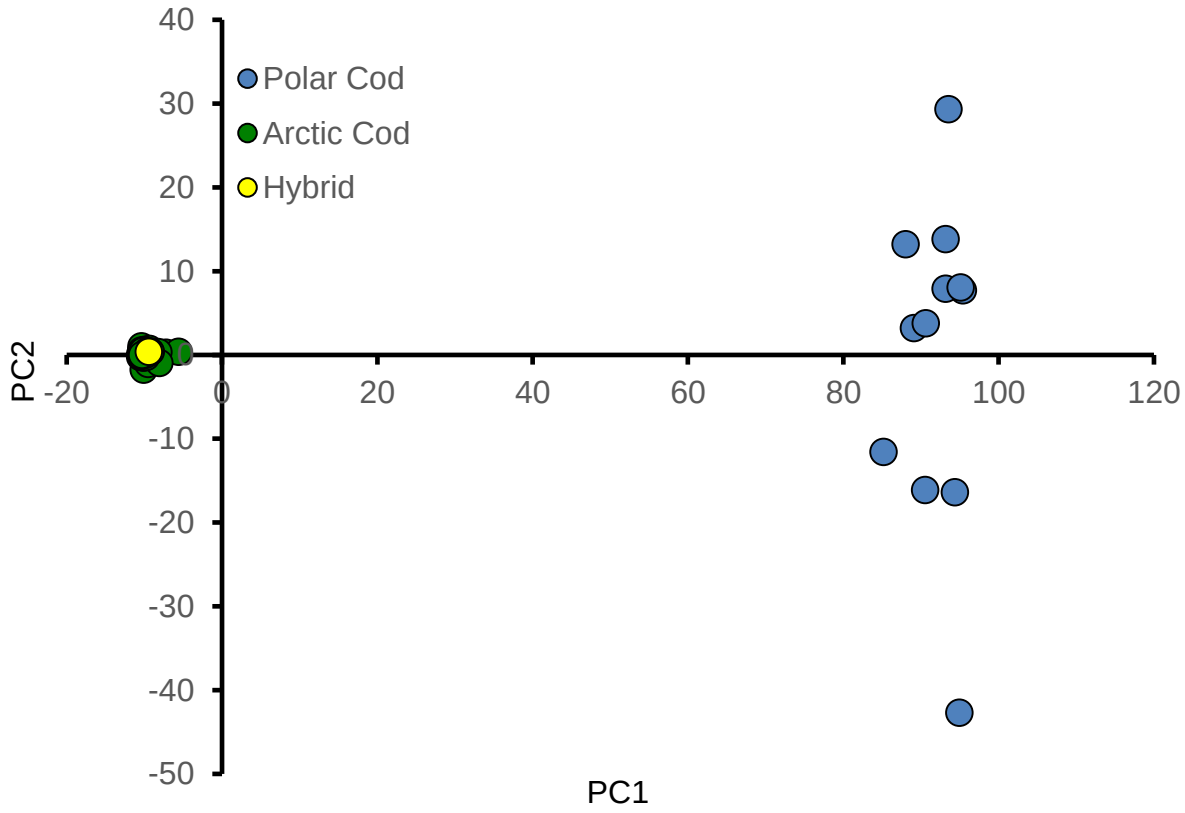


Figure 2-14. Scatterplot of first two principal components based on 847 loci for double digest RAD-Sequencing. Yellow circle indicates a hybrid individual based on possessing a Polar cod mitogenome and a nuclear genome clustering with Arctic cod.

Table 2-1. Nuclear microsatellite DNA primer name, sequence, number of alleles, allele spread, observed heterozygosity (HO), expected heterozygosity (HE), source and USGS Alaska Science Center (ASC) primer redesign information for the loci used in study of Arctic cod genetic diversity data derived from 740 samples.

Primer Name	Universal Tail	Primer Sequence (5' - 3')	No. Alleles	Allele Spread	H_O	H_E	Source	Redesigned By USGS ASC
			14	193-220	0.704	0.697	Nelson et al.	Y
Bsa6-1(F)	M13F†	GCACTAAAGCATGATGAAAGC					2013	Y
Bsa6-1(R)		GACACTGCTTTTATACAGC						
			13	174-198	0.290	0.295	Nelson et al.	Y
Bsa7-1(F)	M13R*	AGGTTCGGTGTTTACCTGC					2013	Y
Bsa7-1(R)		ACTGAACAGGTGTTCAGGC						
			20	180-222	0.602	0.599	Nelson et al.	N
Bsa14(F)	SP6‡	ATGAAATGCTATCCGACTCC					2013	Y
Bsa14-1(R)		GGAAGTCGACTTTTCATGGAC						
			7	179-207	0.558	0.519	Nelson et al.	Y
Bsa15-1(F)	M13F†	ACGCAGTTGGTCCAAAGC					2013	Y
Bsa15-1(R)		ACTCAGAGCTCCTGTTGC						
			13	173-229	0.655	0.669	Nelson et al.	Y
Bsa60-1(F)	M13R*	AATTGAGATTCCCTGGGC					2013	Y
Bsa60-1(R)		ATTTCTGACGTTTCTTGC						
			12	128-158	0.752	0.771	Nelson et al.	N
Bsa101(F)	SP6‡	GTGCTTGTGTGTGTTTCAGC					2013	N
Bsa101(R)		TGTTAATGCTGCTTCTTTGC						
			10	115-151	0.123	0.137	Miller et al.	N
Gmo8(F)	M13F†	TGGGGGAGGCATCTGTCATTCA					2000	N
Gmo8(R)		GCAAAACGAGATGCACAGACACC						
			13	63-99	0.634	0.638	Miller et al.	N
Gmo34(F)	M13R*	GGTTGGACCTCATGGTGAA					2000	N
Gmo34(R)		TCCACAGAAGGTCTCCTAA						
			5	103-115	0.138	0.160	Jakobsdottir	N
PGmo32(F)	SP6‡	CAATCGCCGTCCAACCAAC					et al. 2006	N
PGmo32(R)		GGCGGCAGCAACGATTCTC						
			9	276-316	0.604	0.650	Skirnisdottir	Y
PGmo127-1(F)	SP6‡	CCCTCAAAATTCAACCTGG					et al. 2008	Y
PGmo127-1(R)		TCTCCTCCCCTTGTGTGC						

			31	112-244	0.930	0.938	O'Reilly et.al. 2000	N N	
Tch14(F)	M13F†	CATACATTGGTCACTCTTTCTTAC							
Tch14(R)		AAACTGATATACGCCCAACT							

*M13R sequence: GGATAACAATTTCACACAGG

†M13F sequence: CACGACGTTGTAAACGAC

‡SP6 Promoter sequence: GATTTAGGTGACACTATAG

Table 2-2. Estimates of genetic diversity in Arctic cod in the Beaufort and Chukchi seas and St. Lawrence, Canada, estimated from 11 nuclear microsatellite loci and mitochondrial DNA cytochrome b. Values microsatellite loci include: average number of alleles, allelic richness based on 17 individuals, number of private alleles, observed and expected heterozygosities (H_O/H_E). Values for mitochondrial DNA cytochrome b: number of haplotypes, haplotype richness based on 30 individuals, haplotype diversity (*h*), nucleotide diversity (π). Sample sizes (*N*, *n*) for each marker type are shown. Bold values are significant at $\alpha = 0.05$, Bonferroni corrections applied for microsatellite data. Bold text signifies significant deviations from null expectations.

	Southern Chukchi*	Eastern Chukchi	Western Beaufort	Central Beaufort	Southern Beaufort		St. Lawrence, Canada
					Camden Bay	Southern Shelf	
Microsatellites							
No. Alleles	6.5 (4.2)	7.5 (4.6)	6.5 (4.4)	6.2 (5.0)	10.2 (6.9)	11.6 (7.3)	–
Allelic Richness	6.3 (4.0)	6.0 (3.5)	5.3 (3.4)	5.2 (3.8)	5.4 (3.7)	5.7 (3.7)	–
Private Alleles	2	5	2	2	9	20	–
H_O (SD)	52.4 (3.6)	**51.6 (2.5)**	54.9 (2.5)	54.2 (2.6)	54.2 (1.0)	55.4 (0.8)	–
H_E (SD)	56.6 (7.4)	**57.8 (7.0)**	53.2 (8.2)	54.1 (7.8)	54.4 (7.7)	55.7 (7.7)	–
N	18	37	38	35	236	350	–
mtDNA							
No. Haplotypes	10	–	26	19	36	21	11
Haplotypic Richness	8.2	–	15.0	11.4	12.5	10.4	10.0
h	0.83 (0.04)	–	0.92 (0.02)	0.86 (0.03)	0.88 (0.02)	0.86 (0.02)	0.85 (0.04)
π	0.0041 (0.0024)	–	0.0050 (0.0029)	0.0038 (0.0023)	0.0046 (0.0026)	0.0052 (0.0030)	0.0045 (0.0026)
n	36	–	62	55	117	71	30

* Samples from the Chukchi Sea are pooled for mtDNA analysis due to low samples with known exact locality. Therefore the mtDNA diversity values represent the entire Alaskan Chukchi Sea which includes samples with known (eastern, N = 2 and southern, N = 4) and unknown exact location within the Chukchi Sea (N = 30).

Table 2-3. Observed (H_O) and expected (H_E) heterozygosity statistics for each sampling region, by microsatellite locus. Bold text indicates that region is out of Hardy-Weinberg Equilibrium ($P_{uncorrected} < 0.05$).

Locus	Southern Chukchi		Eastern Chukchi		Chukchi Sea Unknown[1]		Western Beaufort		Central Beaufort		Southern Beaufort			
											Camden Bay		Southern Shelf	
	H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E
Bsa6-1	0.61	0.69	**0.68**	**0.71**	0.69	0.69	0.71	0.70	0.62	0.66	0.74	0.69	0.70	0.70
Bsa7-1	0.39	0.44	**0.27**	**0.31**	0.35	0.31	0.18	0.17	0.27	0.25	0.26	0.27	0.32	0.32
Bsa14	0.67	0.71	**0.43**	**0.60**	0.46	0.57	0.76	0.61	0.49	0.53	0.63	0.60	0.60	0.60
Bsa15-1	0.61	0.49	0.61	0.52	0.50	0.49	0.58	0.52	0.59	0.53	0.55	0.52	0.55	0.52
Bsa60-1	0.50	0.50	**0.58**	**0.66**	0.54	0.57	0.69	0.65	0.63	0.70	0.64	0.64	0.69	0.70
Bsa101	0.61	0.81	**0.70**	**0.82**	0.73	0.78	0.79	0.76	0.89	0.75	0.77	0.77	0.74	0.77
Gmo8	0.06	0.17	**0.09**	**0.19**	0.19	0.21	0.08	0.08	0.15	0.14	0.12	0.13	0.13	0.14
Gmo34	**0.06**	**0.16**	**0.19**	**0.29**	0.08	0.14	0.18	0.17	0.11	0.16	0.16	0.17	0.12	0.14
PGmo32	**0.67**	**0.64**	**0.54**	**0.70**	0.50	0.63	0.62	0.67	0.69	0.71	0.59	0.62	**0.68**	**0.64**
PGmo127-1	0.71	0.67	0.64	0.64	**0.42**	**0.60**	**0.50**	**0.61**	**0.63**	**0.59**	0.60	0.65	0.62	0.66

Tch12	0.89	0.95	0.95	0.93	0.88	0.94	0.95	0.92	0.91	0.94	0.91	0.94	0.95	0.94

[1] Exact locality within the Chukchi Sea for these samples are not known.

Table 2-4. Estimates of population differentiation and Analysis of Molecular Variance (AMOVA) among Arctic cod sampled from six locales within the Beaufort and Chukchi seas of Alaska. Values provided for nuclear microsatellite data include F_{ST} (Weir and Cockerham 1984), R_{ST} (Slatkin 1995), and χ^2 distribution of alleles (Goudet et al. 1996); values for mtDNA cytochrome b include F_{ST} and Φ_{ST} (Excoffier et al. 1992). Significant comparisons are in bold text ($\alpha = 0.05$; P-value = 0.05/11 = 0.0045 for microsatellite loci).

	Microsatellite			mtDNA	
	F_{ST}	R_{ST}	χ^2	F_{ST}	Φ_{ST}
Southern Chukchi					
& Eastern Chukchi	-0.003	-0.008	16.6	--	--
& Western Beaufort	0.004	-0.009	40.3	--	--
& Central Beaufort	0.010	0.009	**43.7**	--	--
& Southern Beaufort – Camden Bay	0.001	-0.004	**52.6**	--	--
& Southern Beaufort – Southern Shelf	0.003	-0.005	**56.8**	--	--
Eastern Chukchi					
& Western Beaufort	-0.001	-0.002	34.3	--	--
& Central Beaufort	0.003	0.012	37.7	--	--
& Southern Beaufort – Camden Bay	0.003	0.008	∞	--	--
& Southern Beaufort – Southern Shelf	0.002	0.005	∞	--	--
Western Beaufort					
& Central Beaufort	0.005	-0.003	37.7	0.002	0.017
& Southern Beaufort – Camden Bay	0.000	-0.005	32.4	-0.002	0.000
& Southern Beaufort – Southern Shelf	0.002	-0.004	25.2	0.009	-0.003
Central Beaufort					
& Southern Beaufort – Camden Bay	0.002	0.002	34.8	0.009	-0.001

& Southern Beaufort – Southern Shelf	0.001	0.003	19.8	0.021	0.035
Southern Beaufort – Camden Bay					
& Southern Beaufort – Southern Shelf	0.000	-0.001	38.6	0.004	0.007
Overall	0.001	0.000	∞	0.003*	-0.011*

*AMOVA analysis was done with all (including samples without exact locality) Beaufort and Chukchi samples.

Table 2-5. List of 14 nonsynonymous codons encoding 13 fixed amino acid replacements inferred to be under positive selection by TREESAAP (z-score > +3.09), PRIME (P < 0.10), MEME (P < 0.05), and/or Fast Unconstrained Bayesian AppRoximation (PP > 0.9); see the list of all replacements in Appendix A2-8-2-9.

Gene	Codon position in gene	Amino acid				TREESAAP				PRIME		Codon Model
		Arctic cod	Polar cod	Walleye pollock	Saffron cod	α -helical tendencies category	Amount	Equilibrium constant (ionization of COOH) category	Amount	α	Amino acid property	
ND1	**2**	L	T	T	T					-16.86	Volume (Contant)	MEME
ND1	29	I	I	L	I					-7.97	Volume (Atchley)	
ND1	72	A	S	S	A					-3.79	Secondary structure factor (Atchley)	
ND1	317	C	W	C	C			6	0.73			
ND2	74	A	S	S	S							MEME
ND2	190	I	I	I	M			8	-0.92			
ND2	205	I	I	V	V							
CO3	159	T	A	A	A	6	-0.59					
ND3	90	T	A	M	A	6	-0.59					

ND3	90	T	A	M	A	6	0.62		
ND4	37	G	G	G	A	8	0.85		
ND4	395	I	I	V	I			8	0.90
ND5	525	T	-	T	A	6	-0.59		
Cytb	257	I	-	T	T			6	-0.74

3 A Transcriptome Resource for the Arctic Cod

3.1 Abstract

The Arctic cod provides a vital link in Arctic food web and thus is considered an important species for environmental monitoring. RNA-Seq was conducted on wild samples representing various age classes and tissue types to obtain as complete a transcriptome as possible on an Illumina MiSeq, which resulted in the total of 64,457 transcripts with an average length of 295 bp. We identified well-known genes that are associated with temperature change or reponse to pollutants. This RNA-Seq effort provides the first insight in the Arctic cod transcriptome, which can be a starting point for investigations identifying genes for local adaptation and genomic responses to future environmental change.

3.2 Introduction

The Arctic cod has been identified as a key component to Arctic food webs as it provides a vital link between lower and higher trophic levels (Bradstreet et al. 1986, Welch et al. 1992, Crawford and Jorgenson 1996). Due to its strong association with sea-ice during certain stages in its life cycle, the Alaska Department of Fish and Game (2015) identified the Arctic cod as a "species of gretest conservation need" because of its position as an ecologically important species. Given the Arctic cod's key placement within the food web as the primary forage fish for upper trophic predators, broad ecological (e.g. climate change) and localized anthropogenic (e.g. pollution) stressors to Arctic cod populations could have impact to other components of the marine ecosystem. The Arctic cod appears highly sensitive to oil pollution (Nahrgang et al. 2016) and therefore may have limited resilience to environmental disturbances (Nahrgang et al. 2014, 2016, Laurel et al. 2016, Drost et al. 2016). Only recently has the genetic makeup of this species received attention; however, genomic resources, which could be used for environmental physiology and population structure studies, are still scarce. Responses to environmental stressors, in particular pollutants, have been shown to be large and diverse, potentially involving thousands of genes (Garcia et al. 2012, Anderson et al. 2015). Thus, a genomic approach is needed to better understand the response of the species to environmental challenges. Currently, researchers are reliant on the Atlantic cod (*Gadus morhua*) genome as a reference for Arctic cod genomics. To provide a starting point for future genomic investigations, we sequenced a partial transcriptome of Arctic cod, using samples collected from differing depths and age classes, and from various tissue types to obtain as complete a transciptome as possible.

3.3 Methods

3.3.1 Sample Collection and Sequencing

Arctic cod were collected at various maximum haul depths (20-1000 m) in the southern Alaskan Beaufort Sea and adjacent Canadian waters as part of the BOEM Central Beaufort Sea Survey and US/Canada Transboundary Survey conducted between August 13 and August 30, 2013 (see Figure 2-1 for general location). To obtain the field-acclimatized transcriptome profile, six whole individuals were preserved in RNAlater® (Sigma-Aldrich, St. Louis, MO, USA) for a maximum of 12 hours, then stored at minus 80°C until processing. We classified three samples as juveniles (i.e. age 0, hatched that year) based on a total length between 21 and 37 mm. We classified the other three specimens as older juveniles (either age 0, 1, or 2) and were of length between 74-103 mm (Fey and Weslawski 2017, B. Norcross, University of Alaska Fairbanks personal communication; Table S-11).

For the three older juveniles, we dissected three tissue types (gill, intestine, and muscle) prior to extraction, while a cross section behind the pectoral fin was taken from age 0 samples encompassing multiple organs. We extracted total RNA from each tissue type (nine total extractions representing 3 samples) and pooled organ (3 samples) using a MasterPure™ RNA Purification Kit (Epicentre, Madison, WI, USA) following manufacturer's protocol. Average RNA yield was 188.5 ng/uL, quantified using a Quant-iT RNA Assay Kit (Invitrogen, Carlsbad, CA USA). We generated twelve libraries from high-quality RNA using the TruSeq® stranded mRNA library preparation kit following the low sample protocol (Illumina, San Diego, CA USA). Concentrations for each final library were quantified using Quant-iT RNA Assay Kit and sequenced on the Illumina MiSeq Desktop Sequencer [2 × 150 bp (base pair) read-length configuration] following the manufacturer's protocol.

3.3.2 De Novo Assembly and Functional Annotation

Transcriptome sequencing of the 12 libraries yielded 13,880,362 150-bp paired end reads with an average of 1.16 (± 0.43) million for each library. Prior to de novo assembly, adaptor trimming and demultiplexing was performed using MiSeq Reporter software (Illumina, San Diego, CA USA). We pooled all reads across individuals before using the de novo assembler in CLC Genomics Workbench v.9.5.1 (CLC Bio, Aarhus Denmark) using default settings in order to recover as complete a transcriptome as possible. A total of 64,457 contigs were obtained with an average length of 295 bp (range: 59 – 5,024 bp) and 47.7% GC content.

Transcripts were annotated using BLAST-X and InterProScan, implemented in Blast2Go (Conesa et al. 2005). A BLAST-X homology search, using 1e-3 E-value cutoff against the NCBI non-redundant (nr) protein database for bony fish (taxa ID: 7898), resulted in a total of 23,714 transcript annotations of which 14,912 were mapped to existing gene categories in Blast2Go. The InterProScan analysis was merged with the mapped gene ontologies resulting in 15,474 annotated contigs (SRA accession: SRP116959: see https://trace.ddbj.nig.ac.jp/DRASearch/study?acc=SRP116959).

3.4 Results and Discussion

Characteristics of the Arctic cod transcriptome sequencing project are shown in Table 3-1. Comparison of the top level gene ontology terms identified 16,426 sequences for "Molecular Function", 39,645 sequences for "Biological Processes", and 31,062 sequences for "Cellular Component" (Figure 3-1).

The main goal of this project was to develop a resource for future Arctic marine ecology research for quantifying the response of Arctic cod to environmental stressors related to climate change (e.g. changing ocean temperatures) or anthropogenic disturbances (e.g. oil exposure). We identified several contigs that have been associated in other species with response to exposure to pollutants or other environmental stressors (see Wilson et al. 2018 for complete list of genes). Thus, these contigs could be used to design markers for reduced representational gene expression panels to assess the impact of environmental perturbations, such as oil spills, introduction of novel pathogens, or environmental changes that may disrupt physiological processes, such as relatively extreme temperature changes.

Our annotation results indicated 321 transcripts were involved with the immune system, 1902 were involved with response to external environmental stimuli, and 45 associated with antioxidant activity. These transcripts included well-known genes that are associated with temperature changes (e.g. heat shock proteins) or response to pollutants, including Cytochrome P450 genes, AHR pathway genes, MHC Class I genes, and interleukin genes (Garcia et al. 2012, Noël et al. 2014, Rosland 2014, Anderson et al. 2015). The antifreeze glycoprotein (AFGP) gene, presumably vital for this species and other polar dwelling fish to survive in sub-zero Celsius water temperatures (Chen et al. 1997), was not identified in the BLAST-X homology search, but a BLAST-n (-discontiguous megablast) search for AFGP identified

496 contigs from the approximately 200kbp locus (GenBank accession JN828577; Zhuang et al. 2012). Unfortunately, as this gene comprises extensive repetitive DNA, which is difficult to analyze using downstream processes such as classical genetic sequencing or next generation resequencing, further development is needed before being able to take full advantage of these AFGP transcriptomic data.

We did not uncover CD4 protein or MHC Class II genes in the partial transcriptome, an interesting result given those same genes are apparently missing from the Atlantic cod genome (Star et al. 2011). Unlike the MHC Class I genes, which are found in both the Arctic and Atlantic cod and which trigger immune response to intracellular pathogens such as viruses, MHC Class II genes present fragments of bacteria and other extracellular pathogens to cells in the immune system, triggering an immune response. In other vertebrates, including other fish species, the proteins CD4 interact with MHC Class II genes and other genes in the MHC Class II pathway involved in making and transporting the MHC Class II proteins. Researchers that reported the lack of MHC Class II and CD4 genes in the genome of the Atlantic cod and transcriptome of another gadoid species, the burbot, posited that cod species inhabiting deep waters may have evolved an immune system that is adapted to a very particular set of pathogens (Star et al. 2011, Star and Jentoff 2012). The consequences of the loss of these genes in the Atlantic cod and, apparently, Arctic cod and other gadid species (Malmstrøm et. al 2016) are unclear. However, it has long been demonstrated that in mice, an MHC Class II-deficient immune system leads to a strong decrease in resistance to pathogens; mice engineered to lack MHC Class II genes suffer severe illnesses (Grusby and Glimcher 1995). It is also known that Atlantic cod respond poorly in terms of antibody production to bacterial infections (Pilstrom et al. 2005). Given the likelihood that climate change in the Arctic may promote the colonization of Arctic waters by novel extracellular pathogens (Hoegh-Guldberg and Bruno 2010), clarification of the ability of the immune system of Arctic cod to novel immune challenges is important.

In addition to comparing the Arctic cod transcriptome reads to the bony fish protein database, we compared the Arctic cod transcripts with the Atlantic cod transcriptome. Based on a local BLAST-n search using Geneious v.10.1.3 (Biomatters Ltd.) and leveraging the predicted transcripts (83,505 predicted proteins/transcripts; Tørresen et al. 2017) as a reference, we found that 68.3% of the Arctic cod contigs (44,002) aligned (i.e. contained a blast hit) using a max E-value of 1e-10 and 85.7% (55,244) with max E-value of 1e-1. Further, pairwise identity ranged from 64.4% to 100% with 32,657 (out of 44,002) transcripts containing at least 90% identity with the predicted Atlantic cod transcriptome. While 74% of the identified Arctic cod transcripts were nearly identical to the closely-related Atlantic cod, a significant number of transcripts demonstrated substantial differences between the two gadid species, confirming a need for a more complete Arctic cod reference transcriptome. Thus, the partial transcriptome presented in this study will provide a key foundation to the identification of genes important for local adaptation and genomic responses to future environmental change, as well as for population differentiation.

All data belong to BioProject PRJNA399761, which can be accessed publicly here: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA399761. Raw reads for each individual library were deposited in the NCBI Sequence Read Archive (SRA accession: SRP116959: see https://trace.ddbj.nig.ac.jp/DRASearch/study?acc=SRP116959). Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GFXU00000000.

3.6 References

Alaska Department of Fish and Game, 2015. Alaska wildlife action plan. Juneau, AK.

Anderson, Ø., M. Frantzen, M. Rosland, G. Timmerhaus, A. Skugor, and A. Krasnov. 2015. Effects of crude oil exposure and elevated temperature on the liver transcriptome of polar cod (*Boreogadus saida*). Aquatic Toxicology 165:9–18.

Bradstreet, M.S.W., K. J. Finley, A. D. Sekarak, W. B. Griffiths, C. R. Evans, M. F. Fabjan, and H. E. Stallard. 1986. Aspects of the biology of Arctic cod (*Boreogadus saida*) and its importance in Arctic marine food chains. Canadian Technical Report of Fisheries and Aquatic Sciences No. 1491. 193 pages.

Chen, L., A.L. DeVries, and C.-H.C. Cheng. 1997. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. Proceedings of the National Academy of Sciences USA 94: 3817–3822.

Conesa, A., S. Gotz, J. M. García-Gómez, J. Teroll, M. Talón, and M. Robles. 2005. Blast2Go: A universal tool for annotation, visualization and analysis for functional genomics research. Bioinformatics 21:3674–3676.

Crawford, R. E., and J. K Jorgenson. 1996. Quantitative studies of Arctic cod, *Boreogadus saida* schools: Important energy stores in the Arctic food web. Arctic 49:181–193.

Drost, H. E., M. Lo, E.C. Carmack, and A.P. Farrell. 2016. Acclimation potential of Arctic cod (*Boreogadus saida* Lepechin) from the rapidly warming Arctic Ocean. Journal of Experimental Biology: doi: 10.1242/jeb.140194.

Fey, D.P. and J.M. Weslawski. 2017. Age, growth rate, and otolith growth of polar cod (*Boreogadus saida*) in two fjords of Svalbard, Kongsfjorden and Rijpfjorden. Oceanologia 59: 576–584.

Garcia, T.I., Y. Shen, D. Crawford, M.F. Oleksiak, A. Whitehead, and R.B. Walter. 2012. RNA-Seq reveals complex genetic response to deepwater horizon oil release in *Fundulus grandis*. BMC Genomics 13:474.

Grusby, M. J., and L. H. Glimcher. 1995. Immune responses in MHC Class II-deficient mice. Annual Review of Immunology 13:417–435.

Hoegh-Guldberg, O. and J.F. Bruno. 2010. The impact of climate change on the world's marine ecosystems. Science 328:1523–1528.

Laurel, B.J., M. Spencer, P. Iseri, and L.A. Copeman. 2016. Temperature-dependent growth and behavior of juvenile Arctic cod (*Boreogadus saida*) and co-occurring North Pacific gadids. Polar Biology 39:1127–1135.

Malmstrøm, M. M. Matchiner, O.K. Tørresen, B. Star, L.G. Snipen, T.F. Hansen, H.T. Baalsrud, A.J. Nederbragt, R. Hanel, W. Salzburger, N.C. Stenseth, K.S. Jakobsen, and S. Jentoft. 2016. Evolution of immune system influences speciation rates in teleost fishes. Nature Genetics 48:1204–1210.

Nahrgang, J., Ø. Varpe, E. Korshunova, S. Murzina, I. G. Hallanger, I. Vieweg, and J. Berge. 2014. Gender specific reproductive strategies of an Arctic key species (*Boreogadus saida*) and implications of climate change. PLoS One 9:e98452.

Nahrgang, J., P. Dubourg, M. Frantzen, D. Storch, F. Dahlke, and J. P. Meador. 2016. Early life stages of an arctic keystone species (*Boreogadus saida*) show high sensitivity to a water-soluble fraction of crude oil. Environmental Pollution 218:605–614.

Noël, M., L. L. Loseto, C. C. Helbing, N. Veldhoen, N.J. Dangerfield, and P. S. Ross. 2014. PCBs are associated with altered gene transcripts profiles in arctic beluga whales (*Delphinapterus leucas*). Environmental Science and Technology 48:2942–2951.

Pilstrom, L., G. Warr, and S. Stromberg. 2005. Why is the antibody response of Atlantic cod so poor? The search for a genetic explanation. Fisheries Science 71:961–971.

Rosland, M. 2014. Effects of oil exposure and elevated temperature on the liver transcriptome in polar cod (*Boreogadus saida).* MS Thesis. Norwegian University of Life Sciences, Ås, Norway.

Star, B.S., A.J. Nederbragt, S. Jentoft, U. Grimholt, M. Malmstrom, T.F. Gregers, et al. 2011. The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210.

Star, B. and S. Jentoft S. 2012. Why does the immune system of Atlantic cod lack MHC II? Bioessays 34:648–651.

Tørresen, O.K., B. Star, S. Jentoft, S., W.B. Reinar, H. Grove, J.R. Miller, B.P. Walenz, J. Knight, J.M. Ekholm, P. Peluso, R.B. Edvardsen, A. Tooming-Klunderud, M. Skage, S. Lien, K.S. Jakobsen, and A.J. Nederbragt. 2017. An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. BMC Genetics 18:95.

Welch, H.E., M.A. Bergmann, T.D. Siferd, K.A. Martin, M.F. Curtis, R.E. Crawford, R.J. Conover, and H. Hop. 1992. Energy flow through the marine ecosystem of the Lancaster Sound region, Arctic Canada. Arctic 45:343–357.

Wilson, R.E., B.J. Pierson, G.K. Sage, M.C. Gravley, D.M. Menning. S.A. Sonsthagen, J. Nelson, and S.L. Talbot. 2018. Genetic and Genomic Data on Arctic cod (*Boreogadus saida*) and other gadids, Beaufort Sea, Alaska, 2011-2014. U.S. Geological Survey data release. http://doi.org/10.5066/F7CF9P23.

Zhuang, X., C. Yang, S. E. Fevolden, and C. Cheng. 2012. Protein genes in repetitive sequence - antifreeze glycopropteins in Atlantic cod genome. BMC Genomics 13:293.

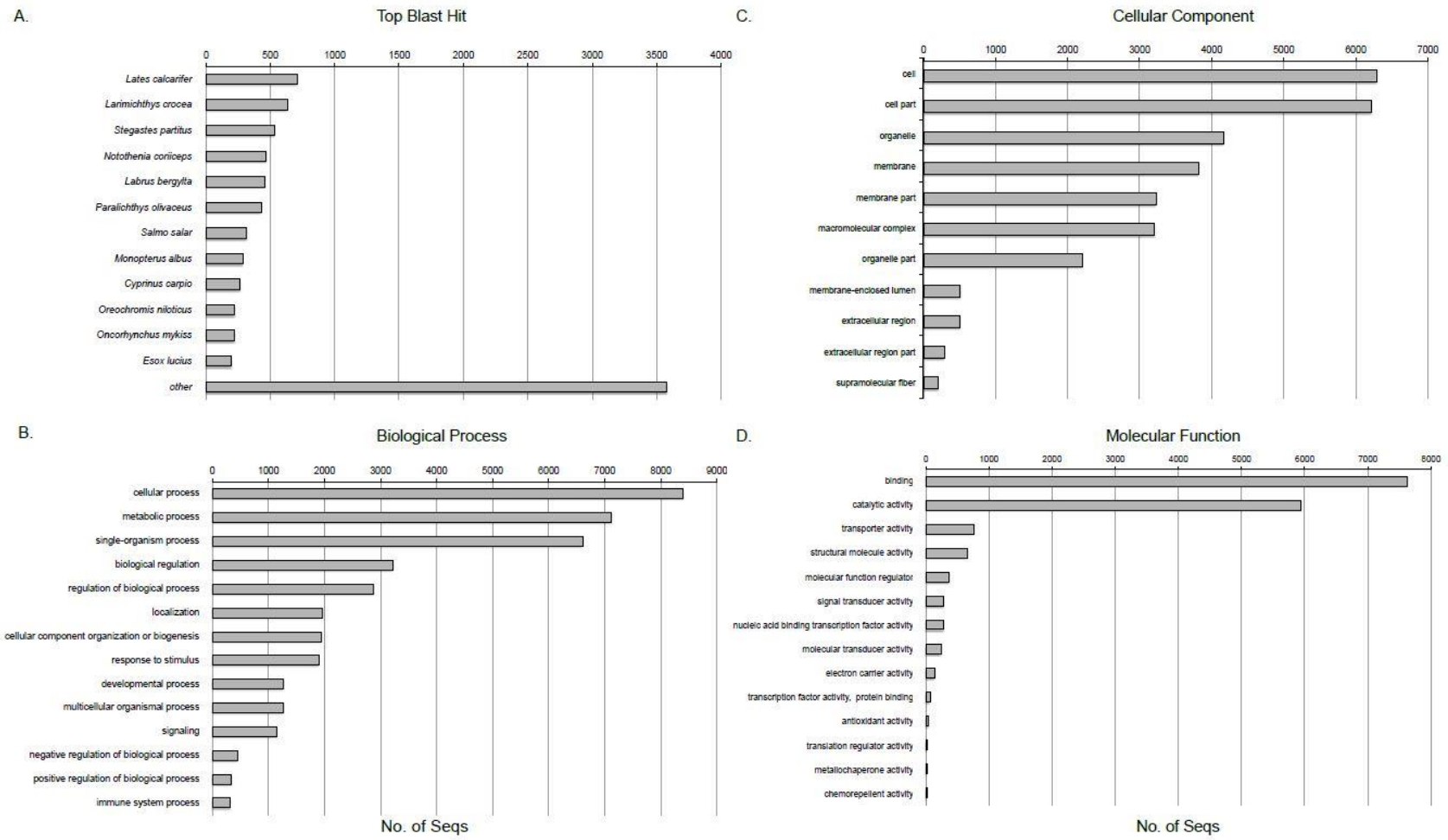


Figure 3-1. Species distribution of top BLAST-X hit (A) and functional profiling of *Boreogadus saida* transcripts using Blast2Go analysis (B-D).

Table S-1. Primers used in this study to amplify the mitochondrial cytochrome b gene for Arctic cod.

Primer	Sequence 5’ to 3’	Orientation	Source
CB115F	TTTGGCTCTCTTCTAGGC	Forward	This study
CB993R	GAATGTTAAACCTCGCTGC	Reverse	This study
BOSA_cybF	ATGCCATAATTCCTGCCCAGA	Forward	This study
BOSA_cybR	CGGTAATAATTGGATGYAGCAC	Reverse	This study

Table S-2. Mitogenome information. Sequences for primers used in Long-Range PCR reactions. Bases within the alignment refer to the alignment of Arctic cod mitogenomes.

PCR fragment	Primer Name	Direction	Primer Sequence (5' - 3')	Location in gene of primer	Bases within alignment
1	BOSA_20L	Forward	TTACACATGCAAGTCTCCGC	12S RNA	110-129
	BOSA_3924H	Reverse	GAAAGTGGTGTAGTGGAAGC	tRNA-Gln	3925-3944
2 & 3	BOSA_2866L	Forward	CGTGGTTAGTTATATCCTCAACCCGC	ND1	2842-2860
	BOSA_5072L	Forward	GGGGCTTAGGATAAACTAGAC	tRNA-Trp	5062-5082
	BOSA_5547H	Reverse	GTGTTACCACGTCAGATTGC	tRNA-Cys	5541-5560
	BOSA_7060H	Reverse	TATTACTTCCCGTTTGGCAGCG	COI	7065-7080
4	BOSA_5816L	Forward	AATGTGATCGTTACAGCGCACG	COI	5800-5821
	BOSA_8966H	Reverse	GGGTCTACTATGTGGTATGC	COIII	8971-8900
5 & 6	BOSA_8277L	Forward	ACTGACCATGACACTAAGCC	ATPase6	8263-8282
	BOSA_10449L	Forward	CTGGACTAGCCCTACTTGT	ND4L	10443-10461
	BOSA_11095H	Reverse	CAGCYCATCAGATTTTACTACC	ND4	11089-11110
	BOSA_13339H	Reverse	AGTGTTAGGGTAAGGGCTCAG	ND5	13344-13364
7	BOSA_12871L	Forward	GTGTCTGCCCTACTTCATTCTAGC	ND5	12853-12876
	BOSA_420H	Reverse	TGTTCTTCGTGACTTCGGATGC	12S RNA	420-441
	BOSA_14447H	Reverse	CCTGGTTAAAATCTGGGCAG	tRNA-Glu	14470-14489

Table S-3. Primers used Sanger sequencing of the mitochondrial tRNA-Pro in Saffron cod and cytochrome b in Arctic cod.

Primer Name	Direction	Universal Tail[1]	Species	Direction	Primer Sequence (5' - 3')
tRNA-ProF	Forward	M13R	Saffron Cod	Forward	GGACAAGTGGCATCAGTAT
tRNA-ProR	Reverse	SP6	Saffron Cod	Reverse	TTGTTAGTGGTATTCAGAGG
CB115F	Forward		Arctic Cod	Forward	TTTGGCTCTCTTCTAGGC
CB993R	Reverse		Arctic Cod	Reverse	GAATGTTAAACCTCGCTGC

[1]M13R: GGATAACAATTTCACACAGG;

SP6: GATTTAGGTGACACTATAG

Table S-4. Results of PartitionFinder. Partitions of sites and loci with different substitution parameters used for the Bayesian phylogenetic analysis of complete mitogenome sequences, using MrBayes. ND6rc refers to reversed and complemented sequence of ND6 gene.

Partition number	Best model	Subset partitions
1	K80+I	ATPase8_1stpos, ND4L_1stpos, ND4_1stpos, ATPase6_1stpos, Cytb_1stpos, ND1_1stpos, ND3_1stpos, ND2_1stpos, ND5_1stpos
2	HKY+I	COII_2ndpos, COI_2ndpos, COIII_2ndpos, ND6rc_2ndpos, ATPase8_2ndpos, ND4L_2ndpos, ND2_2ndpos, ATPase6_2ndpos, Cytb_2ndpos, ND3_2ndpos, ND1_2ndpos, ND5_2ndpos, ND4_2ndpos
3	GTR+G	Cytb_3rdpos, ND5_3rdpos, ND2_3rdpos, ND1_3rdpos, ND4_3rdpos, ND3_3rdpos, COI_3rdpos, COIII_3rdpos, ATPase6_3rdpos, COII_3rdpos, ATPase8_3rdpos, ND4L_3rdpos
4	K80+I	COII_1stpos, COIII_1stpos, COI_1stpos
5	HKY+I	ND6rc_1stpos
6	HKY+G	ND6rc_3rdpos

Table S-5. Genetic diversity per gene for Arctic cod and other gadids. Genes are listed in order of appearance in mitogenome; n- sample size; H- number of haplotypes; Hd- haplotype diversity; π- nucleotide diversity; Td- Tajima's D; Fs- Fu's F. Significant values for Tajima's D and Fu's F are in bold. Protein-coding gene regions are shaded grey.

	Arctic Cod						Polar Cod						Walleye pollock						Saffron Cod					
Domain	n	H	Hd	π	Td	Fs	n	H	Hd	π	Td	Fs	n	H	Hd	π	Td	Fs	n	H	Hd	π	Td	Fs
tRNA-Phe	63	2	0.03	0.0005			11	1	0.00	0.0000			28	1	0.00	0.0000			57	3	0.07	0.0016		
12SrRNA	61	8	0.22	0.0003	**-2.0**	**-15.4**	11	5	0.62	0.0015	**-1.7**	0.7	22	9	0.75	0.0016	-1.7	-5.2	57	10	0.29	0.0003	**-2.3**	**-14.6**
tRNA-Val	61	2	0.03	0.0005			16	4	0.13	0.0017			28	7	0.00	0.0000			57	19	0.07	0.0010		
16SrRNA	61	18	0.48	0.0004	**-2.6**	**-25.2**	16	4	0.52	0.0004	-1.3	-1.0	22	7	0.54	0.0005	-2.1	-4.2	57	19	0.77	0.0007	**-2.1**	**-16.6**
tRNA-Leu2	61	2	0.03	0.0004			16	1	0.00	0.0000			22	1	0.00	0.0000			57	1	0.00	0.0000		
ND1	63	29	0.81	0.0036	**-1.9**	**-23.5**	19	9	0.81	0.0015	-1.5	**-4.9**	26	12	0.79	0.0022	**-1.9**	**-8.7**	57	16	0.81	0.0027	**-1.7**	**-5.4**
tRNA-Ile	63	2	0.03	0.0005			19	1	0.11	0.0015			22	2	0.25	0.0035			57	1	0.00	0.0000		
tRNA-Gln	63	1	0.00	0.0000			19	1	0.00	0.0000			29	1	0.00	0.0000			57	1	0.00	0.0000		
tRNA-Met	63	3	0.06	0.0009			19	1	0.00	0.0000			29	1	0.00	0.0000			57	1	0.00	0.0000		
ND2	63	29	0.88	0.0049	-1.3	**-13.3**	19	7	0.77	0.0012	-0.9	**-2.3**	29	19	0.94	0.0029	**-2.1**	**-13.6**	57	18	0.85	0.0022	**-1.7**	**-9.3**
tRNA-Trp	63	2	0.06	0.0009			19	1	0.00	0.0000			29	1	0.00	0.0000			57	3	0.07	0.0010		
tRNA-Ala	63	1	0.00	0.0000			19	1	0.00	0.0000			29	2	0.13	0.0019			57	2	0.04	0.0005		
tRNA-Asn	63	3	0.06	0.0009			19	1	0.00	0.0000			29	1	0.00	0.0000			57	2	0.04	0.0005		
tRNA-Cys	63	1	0.00	0.0000			19	1	0.00	0.0000			29	1	0.00	0.0000			57	1	0.00	0.0000		
tRNA-Tyr	63	2	0.03	0.0005			19	2	0.11	0.0016			29	2	0.07	0.0010			57	3	0.07	0.0011		
CO1	63	30	0.89	0.0040	**-1.7**	**-11.8**	19	10	0.74	0.0009	**-2.0**	**-6.7**	29	23	0.98	0.0030	**-1.8**	**-16.9**	57	22	0.91	0.0013	**-2.1**	**-17.0**
tRNA-Ser2	63	1	0.00	0.0000			19	1	0.00	0.0000			29	1	0.00	0.0000			57	2	0.04	0.0005		
tRNA-Asp	63	4	0.40	0.0061			19	2	0.46	0.0067			29	2	0.13	0.0020			57	2	0.04	0.0005		
CO2	63	18	0.74	0.0022	**-1.2**	**-13.0**	19	5	0.46	0.0012	-1.4	**-1.8**	29	12	0.73	0.0019	**-2.0**	**-8.5**	57	16	0.78	0.0018	**-1.9**	**-12.5**
tRNA-Lys	63	2	0.23	0.0030			19	1	0.00	0.0000			29	2	0.07	0.0019			57	3	0.22	0.0031		
ATPase8	63	4	0.09	0.0006	**-1.7**	**-4.8**	19	2	0.11	0.0006	-1.2	-0.8	29	3	0.53	0.0037	-0.4	0.5	57	3	0.07	0.0004	**-1.5**	**-3.2**
ATPase6	63	17	0.61	0.0021	**-2.1**	**-12.0**	19	6	0.75	0.0017	-0.6	-1.9	29	17	0.89	0.0039	**-2.1**	**-11.2**	57	15	0.82	0.0022	**-1.8**	**-8.9**
CO3	63	29	0.68	0.0025	-1.6	**-10.2**	19	7	0.81	0.0017	-0.8	**-2.7**	29	12	0.66	0.0017	**-2.3**	**-8.4**	57	11	0.69	0.0023	-1.1	-3.0
tRNA-Gly	63	3	0.06	0.0009			19	2	0.11	0.0015			29	1	0.00	0.0000			57	2	0.04	0.0005		
ND3	63	9	0.39	0.0019	**-1.6**	**-5.5**	19	3	0.29	0.0009	-1.1	-1.2	29	6	0.68	0.0029	-0.5	-1.6	57	4	0.47	0.0024	-0.1	0.5

tRNA-Arg	63	3	0.06	0.0009			19	1	0.00	0.0000			29	1	0.00	0.0000			57	1	0.00	0.0000		
ND4L	63	8	0.31	0.0013	**-1.7**	**-5.5**	19	2	0.11	0.0004	-1.2	-0.8	29	4	0.20	0.0007	**-1.7**	**-3.3**	57	5	0.31	0.0011	-1.4	**-3.1**
ND4	63	32	0.82	0.0048	-1.4	**-13.1**	19	9	0.68	0.0015	-1.5	**-3.4**	29	20	0.96	0.0026	**-2.1**	**-13.6**	57	28	0.95	0.0030	-1.4	**-16.2**
tRNA-His	63	1	0.00	0.0000			19	1	0.00	0.0000			29	2	0.07	0.0010			57	1	0.00	0.0000		
tRNA-Ser	63	1	0.00	0.0000			19	1	0.00	0.0000			29	2	0.19	0.0028			57	1	0.00	0.0000		
tRNA-Leu	63	1	0.00	0.0000			19	1	0.00	0.0000			29	2	0.07	0.0009			57	4	0.10	0.0007		
ND5	63	44	0.98	0.0048	**-1.6**	**-24.7**	19	11	0.92	0.0019	-1.4	**-5.6**	29	25	0.99	0.0043	**-1.7**	**-14.7**	57	32	0.96	0.0029	**-1.7**	**-18.8**
ND6	63	17	0.60	0.0022	**-2.1**	**-14.8**							29	14	0.89	0.0041	**-1.6**	**-8.1**	57	17	0.78	0.0035	**-1.8**	**-10.2**
tRNA-Glu	63	1	0.00	0.0000									28	2	0.07	0.0010			57	4	0.14	0.0020		
Cytb	63	29	0.90	0.0035	**-1.7**	**-17.3**							28	21	0.96	0.0039	**-2.1**	**-13.8**	57	22	0.92	0.0027	**-1.6**	**-11.6**
tRNA-Thr	63	3	0.09	0.0013									28	2	0.07	0.0010			57	3	0.32	0.0046		
tRNA-Pro	63	1	0.00	0.0000									28	3	0.14	0.0020			57	2	0.13	0.0019		
Control Region	63	29	0.86	0.0037									28	17	0.93	0.0031			57	19	0.88	0.0027		
Complete mitogenome	61	61	1.00	0.0030									22	20	0.99	0.0023			57	52	0.99	0.0020		

Table S-6. Genetic diversity per gene for Arctic cod. Genes are listed in order of appearance in mitogenome; n- sample size; H- number of haplotypes; Hd- haplotype diversity; π- nucleotide diversity; Td- Tajima's D; Fs- Fu's F. Significant values for Tajima's D and Fu's F are in bold. Protein-coding gene regions are shaded grey.

	Bering Sea						Chukchi Sea						Beaufort Sea (Alaska/Canada)					
Domain	n	H	Hd	π	Td	Fs	n	H	Hd	π	Td	Fs	n	H	Hd	π	Td	Fs
tRNA-Phe	13	1	0	0.00000			20	2	0.1	0.00147			30	1	0	0.00000		
12SrRNA	13	2	0.154	0.00016	-1.15	-0.54	20	4	0.284	0.00042	**-1.87**	**-2.07**	28	4	0.206	0.00023	**-1.73**	**-3.27**
tRNA-Val	13	1	0	0.00000			20	1	0	0.00000			28	2	0.071	0.00099		
16SrRNA	13	3	0.295	0.00018	-1.46	**-2.21**	20	8	0.589	0.00042	**-2.12**	**-6.89**	28	9	0.497	0.00043	**-2.32**	**-7.46**
tRNA-Leu2	13	1	0	0.00000			20	2	0.1	0.00135			28	1	0	0.00000		
ND1	13	7	0.795	0.00239	-1.38	-1.83	20	11	0.805	0.00309	-1.38	**-3.87**	30	16	0.841	0.00433	-1.36	**-8.16**
tRNA-Ile	13	1	0	0.00000			20	1	0	0.00000			30	2	0.067	0.00095		
tRNA-Gln	13	1	0	0.00000			20	1	0	0.00000			30	1	0	0.00000		
tRNA-Met	13	2	0.154	0.00220			20	1	0	0.00000			30	2	0.067	0.00095		
ND2	13	10	0.949	0.00270	-0.83	**-5.43**	20	13	0.884	0.00457	-1.24	-4.01	30	14	0.867	0.00591	-0.45	-1.5
tRNA-Trp	13	1	0	0.00000			20	2	0.1	0.00143			30	2	0.067	0.00095		
tRNA-Ala	13	1	0	0.00000			20	1	0	0.00000			30	1	0	0.00000		
tRNA-Asn	13	2	0.154	0.00211			20	1	0	0.00000			30	2	0.067	0.00091		
tRNA-Cys	13	1	0	0.00000			20	1	0	0.00000			30	1	0	0.00000		
tRNA-Tyr	13	1	0	0.00000			20	2	0.1	0.00149			30	1	0	0.00000		
CO1	13	7	0.833	0.00248	-0.85	-0.43	20	14	0.889	0.00360	-1.14	-4.44	30	17	0.917	0.00481	-0.65	-2.94
tRNA-Ser2	13	1	0	0.00000			20	1	0	0.00000			30	1	0	0.00000		
tRNA-Asp	13	2	0.282	0.00415			20	3	0.416	0.00642			30	3	0.453	0.00693		
CO2	13	6	0.718	0.00171	-1.44	**-2.62**	20	9	0.753	0.00239	-1.44	**-4.16**	30	9	0.747	0.00235	-0.59	-3.03
tRNA-Lys	13	2	0.282	0.00381			20	2	0.268	0.00363			30	2	0.186	0.00252		
ATPase8	13	2	0.154	0.00092	-1.15	-0.53	20	1	0	0.00000	0	0	30	3	0.131	0.00079	**-1.51**	**-2.35**
ATPase6	13	5	0.628	0.00131	-1.57	**2.04**	20	7	0.584	0.00199	-1.81	**-2.45**	30	10	0.634	0.00252	-1.57	**-3.87**
CO3	13	6	0.718	0.00241	-1.36	-1.28	20	6	0.516	0.00193	-1.38	-1.04	30	12	0.772	0.00294	-1.16	**-4.63**
tRNA-Gly	13	1	0	0.00000			20	1	0	0.00000			30	3	0.131	0.00193		

ND3	13	1	0	0.00000	-0.53	-1.61	20	6	0.516	0.00223	**-1.71**	-3.06	30	5	0.446	0.00247	-0.86	-1.02
tRNA-Arg	13	2	0.154	0.00223			20	1	0	0.00000			30	2	0.067	0.00097		
ND4L	13	3	0.295	0.00104	**-1.73**	**-3.32**	20	3	0.279	0.00097	-1.14	-1.21	30	4	0.352	0.00156	-0.91	-1.38
ND4	13	9	0.872	0.00223	-1.51	**-3.43**	20	13	0.853	0.00455	**-1.57**	-2.7	30	16	0.789	0.00602	-0.4	-1.62
tRNA-His	13	1	0	0.00000			20	1	0	0.00000			30	1	0	0.00000		
tRNA-Ser	13	1	0	0.00000			20	1	0	0.00000			30	1	0	0.00000		
tRNA-Leu	13	1	0	0.00000			20	1	0	0.00000			30	1	0	0.00000		
ND5	13	9	0.936	0.00273	-0.76	-1.75	20	17	0.979	0.00458	-1.28	**-6.45**	30	26	0.991	0.00571	-1.15	**-12.44**
ND6	13	4	0.423	0.00118	**-1.77**	**-1.56**	20	8	0.589	0.00202	**-153**	**-4.7**	30	11	0.68	0.00268	**-1.59**	**-6.37**
tRNA-Glu	13	1	0	0.00000			20	1	0	0.00000			30	1	0	0.00000		
Cytb	13	11	0.962	0.00331	-1.13	**-5.9**	20	8	0.826	0.00271	-0.56	-0.75	30	18	0.92	0.00411	-1.12	**-7.26**
tRNA-Thr	13	1	0	0.00000			20	2	0.1	0.00141			30	3	0.131	0.00188		
tRNA-Pro	13	1	0	0.00000			20	1	0	0.00000			30	1	0	0.00000		
Control region	13	8	0.808	0.00271			20	10	0.795	0.00320			30	18	0.903	0.00444		
Complete mitogenome	13	13	1	0.00190			20	20	1	0.00275			30	28	1	0.00369		

Table S-7. Genetic diversity per gene for Walleye pollock. Genes are listed in order of appearance in mitogenome; n- sample size; H- number of haplotypes; Hd- haplotype diversity; π- nucleotide diversity; Td- Tajima's D; Fs- Fu's F. Significant values for Tajima's D and Fu's F are in bold. Protein-coding gene regions are shaded grey.

	Aleutian Islands						Bering Sea						Chukchi/Beaufort Sea					
Domain	n	H	Hd	π	Td	Fs	n	H	Hd	π	Td	Fs	n	H	Hd	π	Td	Fs
tRNA-Phe	12	1	0	0			8	1	0	0			1	3	0	0		
12SrRNA	12	6	0.758	0.00119	**-1.63**	**-4.17**	8	4	0.643	0.00169	-0.76	-1.11	2	2	1	0.0042	0	1.61
tRNA-Val	12	1	0	0			8	1	0	0			1	2	0	0		
16SrRNA	12	4	0.561	0.00058	-1.53	-0.68	8	3	0.464	0.0003	-1.31	-0.99	2	2	1	0.0006	0	0
tRNA-Leu2	12	1	0	0			8	1	0	0			1	2	0	0		
ND1	12	5	0.576	0.00162	-1.62	-0.78	11	9	0.945	0.00291	-1.11	**-4.00**	6	6	1	0.00403	-1.11	-2.66
tRNA-Ile	12	2	0.303	0.00427			11	2	0.182	0.00256			6	2	0.333	0.00469		
tRNA-Gln	12	1	0	0			11	1	0	0			6	1	0	0		
tRNA-Met	12	1	0	0			11	1	0	0			6	1	0	0		
ND2	12	8	0.894	0.00201	**-1.76**	**-3.82**	11	10	0.982	0.00354	-1.62	**-5.79**	6	6	1	0.00401	-1.22	-2.52
tRNA-Trp	12	1	0	0			11	1	0	0			6	1	0	0		
tRNA-Ala	12	2	0.303	0.00439			11	1	0	0			6	1	0	0		
tRNA-Asn	12	1	0	0			11	1	0	0			6	1	0	0		
tRNA-Cys	12	1	0	0			11	1	0	0			6	1	0	0		
tRNA-Tyr	12	1	0	0			11	1	0	0			6	2	0.333	0.00498		
CO1	12	10	0.97	0.00276	-1.03	**-4.25**	11	11	1	0.00347	**-1.7**	**-6.67**	6	6	1	0.00262	-0.94	-2.59
tRNA-Ser2	12	1	0	0			11	1	0	0			6	1	0	0		
tRNA-Asp	12	2	0.303	0.00459			11	1	0	0			6	1	0	0		
CO2	12	5	0.576	0.00119	**-1.83**	**-2.38**	11	8	0.891	0.00229	**-1.37**	**-5.71**	6	4	0.8	0.00286	**-1.71**	-0.5
tRNA-Lys	12	1	0	0			11	1	0	0			6	2	0.333	0.00901		
ATPase8	12	2	0.409	0.00244	0.82	0.73	11	2	0.545	0.00325	1.44	1.13	6	3	0.5	0.00754	-0.19	0.21
ATPase6	12	7	0.894	0.0025	-1.41	**-3.06**	11	8	0.891	0.00362	**-1.72**	**-3.67**	6	5	0.933	0.0076	-1.28	-0.28
CO3	12	5	0.576	0.00148	**-1.94**	-1.51	11	7	0.818	0.00227	**-1.77**	**-3.25**	6	3	0.6	0.00127	-1.23	-0.19
tRNA-Gly	12	1	0	0			11	1	0	0			6	1	0	0		

ND3	12	3	0.53	0.00186	-0.05	-0.14	11	5	0.764	0.00332	-0.54	-1.68	6	5	0.933	0.00494	-0.06	**-2.43**
tRNA-Arg	12	1	0	0			11	1	0	0			6	1	0	0		
ND4L	12	1	0	0	0	0	11	1	0	0	0	0	6	4	0.8	0.00337	-1.23	-1.81
ND4	12	7	0.879	0.00151	-0.84	-2.42	11	11	1	0.00357	-1.69	**-7.07**	6	6	1	0.00327	-1.24	-2.36
tRNA-His	12	1	0	0			11	2	0.182	0.00264			6	1	0	0		
tRNA-Ser	12	2	0.167	0.00245			11	2	0.182	0.00267			6	2	0.333	0.0049		
tRNA-Leu	12	1	0	0			11	2	0.182	0.00249			6	1	0	0		
ND5	12	10	0.97	0.00348	-1.15	-2.77	11	11	1	0.00471	-1.38	**-4.7**	6	6	1	0.00555	-0.66	-0.93
ND6	12	7	0.879	0.00406	-0.79	-2.37	11	7	0.909	0.00313	-0.79	**-3.54**	6	5	0.933	0.00639	**-1.43**	-1.08
tRNA-Glu	12	1	0	0			11	1	0	0			6	1	0	0		
Cytb	12	9	0.939	0.00307	-1.42	-3.36	11	10	0.982	0.0046	**-1.59**	**-4.26**	6	5	1	0.00448	-0.7	-1.28
tRNA-Thr	12	1	0.167	0.00235			11	1	0	0			6	1	0	0		
tRNA-Pro	12	1	0	0			11	1	0	0			6	1	0	0		
Control region	12	8	0.894	0.00279			10	9	0.978	0.00393			5	5	1	0.004		
Complete mitogenome	12	10	0.97	0.00197			8	8	1	0.00283			2	2	1	0.0032		

Table S-8. Genetic diversity per gene for Saffron cod. Genes are listed in order of appearance in mitogenome; n- sample size; Hd- haplotype diversity; π- nucleotide diversity; Td- Tajima's D; Fs- Fu's F. Significant values for Tajima's D and Fu's F are in bold. Protein-coding gene regions are shaded grey.

	Gulf of Alaska						Bering Sea						Chukchi Sea					
Domain	n	H	Hd	π	Td	Fs	n	H	Hd	π	Td	Fs	n	H	Hd	π	Td	Fs
tRNA-Phe	13	1	0	0			21	2	0.095	0.0014			23	2	0.087	0.00256		
12SrRNA	13	2	0.154	0.00016	-1.15	-0.54	21	5	0.352	0.0004	**-1.87**	**-3.74**	23	5	0.324	0.00037	**-1.88**	**-4.89**
tRNA-Val	13	2	0.154	0.00214			21	2	0.095	0.00132			23	1	0	0		
16SrRNA	13	3	0.615	0.00106	1.26	1.16	21	8	0.671	0.00079	**-1.61**	**-3.77**	23	11	0.692	0.00068	**-2.23**	**-7.32**
tRNA-Leu2	13	1	0	0			21	1	0	0			23	1	0	0		
ND1	13	3	0.654	0.00158	-0.16	0.44	21	8	0.719	0.00226	**-1.85**	-1.74	23	10	0.874	0.00354	-1.07	-1.71
tRNA-Ile	13	1	0	0			21	1	1	0			23	1	0	0		
tRNA-Gln	13	1	0	0			21	1	0	0			23	1	0	0		
tRNA-Met	13	1	0	0			21	1	0	0			23	1	0	0		
ND2	13	4	0.679	0.00196	0.97	1.13	21	12	0.871	0.00182	**-1.84**	**-7.77**	23	8	0.802	0.00207	-0.94	-1.59
tRNA-Trp	13	2	0.154	0.0022			21	2	0.095	0.00136			23	1	0	0		
tRNA-Ala	13	1	0	0			21	2	0.095	0.00138			23	1	0	0		
tRNA-Asn	13	1	0	0			21	2	0.095	0.0013			23	1	0	0		
tRNA-Cys	13	1	0	0			21	1	0	0			23	1	0	0		
tRNA-Tyr	13	1	0	0			21	2	0.095	0.00142			23	1	0	0		
CO1	13	3	0.615	0.00104	2.12	1.87	21	11	0.895	0.00104	**-1.82**	**-7.11**	23	14	0.921	0.00142	**-1.77**	**-9.61**
tRNA-Ser2	13	1	0	0			21	2	0.095	0.00134			23	1	0	0		
tRNA-Asp	13	1	0	0			21	1	0	0			23	2	0.087	0.00128		
CO2	13	5	0.731	0.00145	-0.76	-1.75	21	8	0.762	0.00159	**-1.68**	**-4.47**	23	9	0.806	0.00255	-0.82	**-4.44**
tRNA-Lys	13	1	0	0			21	3	0.186	0.00254			23	2	0.087	0.00118		
ATPase8	13	1	0	0	0	0	21	1	0	0	0	0	23	6	0.656	0.00471	**-1.51**	**-2.03**
ATPase6	13	2	0.538	0.00158	1.88	2.66	21	9	0.8	0.00187	**-1.66**	**-5.23**	23	8	0.7	0.00147	-1.19	**-3.28**
CO3	13	4	0.718	0.00242	1.6	0.93	21	4	0.414	0.00147	-1.32	0.38	23	8	0.759	0.00204	-1.13	-2.65
tRNA-Gly	13	1	0	0			21	1	0	0			23	1	0	0		

ND3	13	2	0.538	0.00309	1.88	2.66	21	19	0.986	0.00135	-0.29	0.61	23	3	0.379	0.00116	-1.09	-0.83
tRNA-Arg	13	1	0	0			21	1	0	0			23	3	0.17	0.00252		
ND4L	13	2	0.538	0.00181	1.47	1.23	21	2	0.095	0.00032	-1.16	-0.92	23	5	0.451	0.00194	**-1.51**	**-2.03**
ND4	13	3	0.603	0.00245	1.89	4.22	21	17	0.976	0.00286	-1.22	**-11.67**	23	15	0.96	0.0022	**-1.61**	**-9.34**
tRNA-His	13	1	0	0			21	1	0	0			23	3	0.17	0.00256		
tRNA-Ser	13	1	0	0			21	1	0	0			23	3	0.17	0.00256		
tRNA-Leu	13	1	0	0			21	2	0.095	0.00067			23	2	0.237	0.00325		
ND5	13	5	0.692	0.00201	1.08	-4.02	21	13	0.929	0.00243	-1.26	1.53	23	19	0.98	0.00337	-1.26	**-10.98**
ND6	13	3	0.615	0.00442	1.54	2.88	21	6	0.667	0.00186	-1.31	-2.27	23	8	0.676	0.00224	**-1.84**	**-6.42**
tRNA-Glu	13	1	0	0			21	2	0.095	0.00138			23	2	0.087	0.00126		
Cytb	13	3	0.603	0.00249	1.74	3.58	21	12	0.914	0.0025	**-1.32**	**-5.21**	23	15	0.929	0.00284	**-2.03**	**-8.41**
tRNA-Thr	13	1	0	0			21	3	0.529	0.00791			23	2	0.237	0.00334		
tRNA-Pro	13	1	0	0			21	2	0.181	0.00259			23	1	0	0		
Control region	13	5	0.692	0.00258			21	10	0.857	0.00248			23	11	0.866	0.00263		
Complete mitogenome	13	9	0.872	0.00176			21	21	1	0.00173			23	23	1	0.00197		

Table S-9. List of 122 amino acid replacements among four cod species across 99 codons with non-synomous substitutions with inferences of radical (categories 6-8) changes indicating positive selection (z-score > +3.09) in two amino-acid properties, Alpha-helical tendencies (alpha) and Equilibrium constant (ionization of COOH) (equil); category (cat) and amount (amnt) of change is given, changes that were found to be significant by 10-codon sliding window approach are highlighted in green. Codons with amino acid replacements at multiple nodes are shaded grey. Amino acid replacements inferred to be under selection based on TREESAAP involving fixed differences between species and no variants within species are shaded: yellow- Polar cod, blue- Walleye pollock, orange- Arctic cod, and red- Saffron cod. Hash (#) indicates codons inferred to be under positive selection by MEME ($P < 0.05$) and asterick (*) indicates a codon inferred to be under selection for property change by PRIME ($P < 0.1$); these codons are also in bold.

Gene	Codon No.	From Codon	To Codon	From AA	To AA	Alpha-helical tendencies category	Amnt	Equilibrium constant (ionization of COOH)	Amnt
ND1#*	**2#***	ACC	CTC	Thr	Leu	4	0.38	3	0.26
ND1	3	AAT	AGT	Asn	Ser	1	0.1	2	0.19
ND1*	**29***	ATT	CTT	Ile	Leu	2	0.13	8	1
ND1*	**72***	GCC	TCC	Ala	Ser	6	-0.65	2	-0.13
ND1*	**72***	TCC	GCC	Ser	Ala	6	0.65	2	0.13
ND1	173	ACA	ATA	Thr	Met	6	0.62	2	0.18
ND1	173	ATA	GTA	Met	Val	4	-0.39	1	0.04
ND1	181	GCC	GTC	Ala	Val	4	-0.36	1	-0.02
ND1	182	CTC	ATC	Leu	Ile	2	-0.13	8	-1
ND1	245	GCT	ACT	Ala	Thr	6	-0.59	2	-0.24
ND1	245	GCT	ACT	Ala	Thr	6	-0.59	2	-0.24
ND1	251	TCC	CCC	Ser	Pro	2	-0.2	2	-0.22
ND1	274	ATA	GTA	Met	Val	4	-0.39	1	0.04
ND1	317	TGC	TGA	Cys	Trp	4	0.38	6	0.73

ND2	332	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND2#	**398#**	AGC	GCC	Ser	Ala	6	0.65	2	0.13
ND2	413	GAA	GAC	Glu	Asp	5	-0.5	2	-0.18
ND2	413	GAC	GAA	Asp	Glu	5	0.5	2	0.18
ND2	419	GCT	GGT	Ala	Gly	8	-0.85	1	0
ND2	483	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND2	483	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND2	514	ATC	ATG	Ile	Met	4	0.37	8	0.92
ND2	528	TTA	ATA	Leu	Met	3	0.24	1	-0.08
ND2	529	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ND2	550	GTA	ATA	Val	Met	4	0.39	1	-0.04
ND2	569	ATA	TTA	Met	Leu	3	-0.24	1	0.08
ND2	598	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND2	632	CTT	ATT	Leu	Ile	2	-0.13	8	-1
ND2	641	CAA	CGG	Gln	Arg	2	-0.13	3	-0.36
ND2	643	ACA	GCA	Thr	Ala	6	0.59	2	0.24
ND2	644	GCA	ATA	Ala	Met	1	0.03	1	-0.06
ND2	647	TTT	CTT	Phe	Leu	1	0.08	4	0.47
ND2	647	TTT	TTA	Phe	Leu	1	0.08	4	0.47
ND2	653	GCA	ACA	Ala	Thr	6	-0.59	2	-0.24
ND2	653	GCA	GTA	Ala	Val	4	-0.36	1	-0.02
ND2	663	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
CO1	1159	ATA	TTA	Met	Leu	3	-0.24	1	0.08
CO2	1311	ATC	GTC	Ile	Val	1	-0.02	8	0.96

CO2	1317	GCC	GTC	Ala	Val	4	-0.36	1	-0.02
CO2	1331	GTT	ATT	Val	Ile	1	0.02	8	-0.96
CO2	1379	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ATP8	1447	ATA	ACA	Met	Thr	6	-0.62	2	-0.18
ATP8	1459	GGT	AGC	Gly	Ser	2	0.2	2	-0.13
ATP8	1466	GCC	TCC	Ala	Ser	6	-0.65	2	-0.13
ATP6	1504	ATC	GTC	Ile	Val	1	-0.02	8	0.96
ATP6	1508	ACT	AGT	Thr	Ser	1	-0.06	1	0.11
ATP6	1513	AGC	AAC	Ser	Asn	1	-0.1	2	-0.19
ATP6#	**1538#**	GGA	GAA	Gly	Glu	8	0.94	2	-0.15
ATP6#	**1538#**	GGA	CGA	Gly	Arg	4	0.41	5	-0.53
ATP6	1551	ATG	ATC	Met	Ile	4	-0.37	8	-0.92
ATP6	1562	TTA	ATA	Leu	Met	3	0.24	1	-0.08
ATP6	1614	ATT	GTC	Ile	Val	1	-0.02	8	0.96
ATP6	1661	GCA	ACA	Ala	Thr	6	-0.59	2	-0.24
ATP6	1661	GCA	ACA	Ala	Thr	6	-0.59	2	-0.24
ATP6	1666	ACA	TCG	Thr	Ser	1	-0.06	1	0.11
ATP6	1675	ATA	ACA	Met	Thr	6	-0.62	2	-0.18
CO3	1744	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
CO3	1791	GTC	ATC	Val	Ile	1	0.02	8	-0.96
CO3	1859	GCA	ACA	Ala	Thr	6	-0.59	2	-0.24
CO3	1879	CTA	ATA	Leu	Met	3	0.24	1	-0.08
CO3	1883	GAT	GAA	Asp	Glu	5	0.5	2	0.18
ND3	1977	TTA	TCA	Leu	Ser	4	-0.44	2	-0.15

ND3	1981	CTA	GTA	Leu	Val	2	-0.15	1	-0.04
ND3	2047	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND3	2047	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND3	2047	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND3	2051	GCA	ACA	Ala	Thr	6	-0.59	2	-0.24
ND3	2051	ACA	ATG	Thr	Met	6	0.62	2	0.18
ND3	2051	ATG	GTG	Met	Val	4	-0.39	1	0.04
ND3	2051	ACA	GCA	Thr	Ala	6	0.59	2	0.24
ND4L	2084	ACA	TCA	Thr	Ser	1	-0.06	1	0.11
ND4L	2093	TTA	ATA	Leu	Met	3	0.24	1	-0.08
ND4L	2125	TCC	GCC	Ser	Ala	6	0.65	2	0.13
ND4L	2128	TTA	TCA	Leu	Ser	4	-0.44	2	-0.15
ND4L	2131	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND4L	2135	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND4L	2169	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND4	2212	GGC	GCC	Gly	Ala	8	0.85	1	0
ND4	2215	ACC	TCC	Thr	Ser	1	-0.06	1	0.11
ND4	2223	ACG	ATG	Thr	Met	6	0.62	2	0.18
ND4	2226	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND4	2234	ACA	GCA	Thr	Ala	6	0.59	2	0.24
ND4	2258	AAC	AGC	Asn	Ser	1	0.1	2	0.19
ND4	2276	GCT	ACT	Ala	Thr	6	-0.59	2	-0.24
ND4	2276	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND4	2276	GCT	ACT	Ala	Thr	6	-0.59	2	-0.24

ND4	2296	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ND4	2342	ATG	ACG	Met	Thr	6	-0.62	2	-0.18
ND4	2358	GCT	ACT	Ala	Thr	6	-0.59	2	-0.24
ND4	2377	ATC	GTC	Ile	Val	1	-0.02	8	0.96
ND4	2425	GTC	ATC	Val	Ile	1	0.02	8	-0.96
ND4#	**2449#**	GGC	AGC	Gly	Ser	2	0.2	2	-0.13
ND4	2570	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND4	2574	ATT	GTC	Ile	Val	1	-0.02	8	0.96
ND4	2574	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND4	2614	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND4#	**2618#**	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND4#	**2618#**	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND4#	**2618#**	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND5	2656	TAC	TGC	Tyr	Cys	1	0.01	5	-0.55
ND5	2656	TAT	TGT	Tyr	Cys	1	0.01	5	-0.55
ND5	2668	AAT	AGT	Asn	Ser	1	0.1	2	0.19
ND5	2668	AGT	AAT	Ser	Asn	1	-0.1	2	-0.19
ND5	2671	TGA	TCA	Trp	Ser	3	-0.31	2	-0.17
ND5	2696	AGC	ACC	Ser	Thr	1	0.06	1	-0.11
ND5	2702	GTA	ATA	Val	Met	4	0.39	1	-0.04
ND5	2703	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND5#	**2719#**	AGC	GGC	Ser	Gly	2	-0.2	2	0.13
ND5	2753	CTC	TTC	Leu	Phe	1	-0.08	4	-0.47
ND5	2754	ATA	ATT	Met	Ile	4	-0.37	8	-0.92

ND5	2755	AAT	AGT	Asn	Ser	1	0.1	2	0.19
ND5	2827	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND5	2847	GAT	GGC	Asp	Gly	4	-0.44	3	0.33
ND5	2847	GGC	AGC	Gly	Ser	2	0.2	2	-0.13
ND5	2851	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND5	2852	CTC	TTC	Leu	Phe	1	-0.08	4	-0.47
ND5	2898	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND5	2904	ATA	TTA	Met	Leu	3	-0.24	1	0.08
ND5	2913	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND5	2913	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND5	2916	ACT	TCT	Thr	Ser	1	-0.06	1	0.11
ND5	3012	TCC	ACC	Ser	Thr	1	0.06	1	-0.11

Table S-10. List of 164 amino acid replacements from three cod species across 137 codons with non-synomous substitutions with inferences of radical (categories 6-8) changes indicating positive selection (z-score > +3.09) in two amino-acid properties, Alpha-helical tendencies (alpha) and Equilibrium constant (ionization of COOH) (equil); category (cat) and amount (amnt) of change is given, changes that were found to be significant by 10-codon sliding window approach are highlighted in green. Codons with amino acid replacements at multiple nodes are shaded grey. Amino acid replacements inferred to be under selection based on TREESAAP involving fixed differences between species and no variants within species are shaded: blue- Walleye pollock, orange- Arctic cod, and red- Saffron cod. Hash (#) indicates codons inferred to be under positive selection by MEME ($P < 0.05$) and asterick (*) indicates a codon inferred to be under selection for property change by PRIME ($P < 0.1$); these codons are also in bold.

Gene	Codon#	From Codon	To Codon	From AA	To AA	Alpha-helical tendencies	Amnt	Equilibrium constant (ionization of COOH)	Amnt
ND1#*	**2#***	ACC	CTC	Thr	Leu	4	0.38	3	0.26
ND1	3	AAT	AGT	Asn	Ser	1	0.1	2	0.19
ND1*	**29***	ATT	CTT	Ile	Leu	2	0.13	8	1
ND1*	**72***	GCC	TCC	Ala	Ser	6	-0.65	2	-0.13
ND1	173	ACA	ATA	Thr	Met	6	0.62	2	0.18
ND1	173	ATA	GTA	Met	Val	4	-0.39	1	0.04
ND1	182	CTC	ATC	Leu	Ile	2	-0.13	8	-1
ND1	245	GCT	ACT	Ala	Thr	6	-0.59	2	-0.24
ND1	251	TCC	CCC	Ser	Pro	2	-0.2	2	-0.22
ND1	274	ATA	GTA	Met	Val	4	-0.39	1	0.04
ND2	332	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND2#	**398#**	AGC	GCC	Ser	Ala	6	0.65	2	0.13
ND2	413	GAA	GAC	Glu	Asp	5	-0.5	2	-0.18
ND2	419	GCT	GGT	Ala	Gly	8	-0.85	1	0
ND2	483	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND2	514	ATG	ATC	Met	Ile	4	-0.37	8	-0.92
ND2	529	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ND2	550	GTA	ATA	Val	Met	4	0.39	1	-0.04
ND2	569	TTA	ATA	Leu	Met	3	0.24	1	-0.08
ND2	598	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND2	632	CTT	ATT	Leu	Ile	2	-0.13	8	-1
ND2	641	CGG	CAA	Arg	Gln	2	0.13	3	0.36
ND2	644	ATA	GCA	Met	Ala	1	-0.03	1	0.06
ND2	653	ACA	GCA	Thr	Ala	6	0.59	2	0.24
ND2	653	GCA	GTA	Ala	Val	4	-0.36	1	-0.02
ND2	663	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
CO1	1159	TTA	ATA	Leu	Met	3	0.24	1	-0.08
CO2	1311	ATC	GTC	Ile	Val	1	-0.02	8	0.96
CO2	1317	GCC	GTC	Ala	Val	4	-0.36	1	-0.02
CO2	1331	GTT	ATT	Val	Ile	1	0.02	8	-0.96

CO2	1379	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ATP8	1447	ATA	ACA	Met	Thr	6	-0.62	2	-0.18
ATP8	1459	AGC	GGT	Ser	Gly	2	-0.2	2	0.13
ATP8	1466	TCC	GCC	Ser	Ala	6	0.65	2	0.13
ATP6	1504	ATC	GTC	Ile	Val	1	-0.02	8	0.96
ATP6	1508	ACT	AGT	Thr	Ser	1	-0.06	1	0.11
ATP6	1513	AGC	AAC	Ser	Asn	1	-0.1	2	-0.19
ATP6#	**1538#**	GGA	CGA	Gly	Arg	4	0.41	5	-0.53
ATP6#	**1538#**	GGA	GAA	Gly	Glu	8	0.94	2	-0.15
ATP6	1551	ATG	ATC	Met	Ile	4	-0.37	8	-0.92
ATP6	1562	TTA	ATA	Leu	Met	3	0.24	1	-0.08
ATP6	1614	GTC	ATT	Val	Ile	1	0.02	8	-0.96
ATP6	1661	ACA	GCA	Thr	Ala	6	0.59	2	0.24
ATP6	1666	TCG	ACA	Ser	Thr	1	0.06	1	-0.11
ATP6	1675	ATA	ACA	Met	Thr	6	-0.62	2	-0.18
CO3	1744	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
CO3	1791	GTC	ATC	Val	Ile	1	0.02	8	-0.96
CO3	1859	GCA	ACA	Ala	Thr	6	-0.59	2	-0.24
CO3	1879	ATA	CTA	Met	Leu	3	-0.24	1	0.08
CO3	1883	GAA	GAT	Glu	Asp	5	-0.5	2	-0.18
ND3	1977	TTA	TCA	Leu	Ser	4	-0.44	2	-0.15
ND3	1981	CTA	GTA	Leu	Val	2	-0.15	1	-0.04
ND3	2047	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND3	2051	GCA	ACA	Ala	Thr	6	-0.59	2	-0.24
ND3	2051	ACA	ATG	Thr	Met	6	0.62	2	0.18
ND3	2051	ATG	GTG	Met	Val	4	-0.39	1	0.04
ND4L	2084	ACA	TCA	Thr	Ser	1	-0.06	1	0.11
ND4L	2131	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND4L	2135	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND4L	2169	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND4	2212	GCC	GGC	Ala	Gly	8	-0.85	1	0
ND4	2215	TCC	ACC	Ser	Thr	1	0.06	1	-0.11
ND4	2223	ACG	ATG	Thr	Met	6	0.62	2	0.18
ND4	2226	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND4	2234	ACA	GCA	Thr	Ala	6	0.59	2	0.24
ND4	2258	AAC	AGC	Asn	Ser	1	0.1	2	0.19
ND4	2276	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND4	2276	GCT	ACT	Ala	Thr	6	-0.59	2	-0.24
ND4	2276	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND4	2296	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ND4	2342	ATG	ACG	Met	Thr	6	-0.62	2	-0.18
ND4	2358	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND4	2377	ATC	GTC	Ile	Val	1	-0.02	8	0.96

ND4	2425	ATC	GTG	Ile	Val	1	-0.02	8	0.96
ND4#	**2449#**	GGC	AGC	Gly	Ser	2	0.2	2	-0.13
ND4	2570	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND4	2574	GTC	ATT	Val	Ile	1	0.02	8	-0.96
ND4	2574	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND4	2614	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND4#	**2618#**	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND4#	**2618#**	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND4#	**2618#**	ATC	ACC	Ile	Thr	3	-0.25	6	0.74
ND5	2656	TAT	TGT	Tyr	Cys	1	0.01	5	-0.55
ND5	2656	TAC	TGC	Tyr	Cys	1	0.01	5	-0.55
ND5	2668	AAT	AGT	Asn	Ser	1	0.1	2	0.19
ND5	2668	AGT	AAT	Ser	Asn	1	-0.1	2	-0.19
ND5	2671	TGA	TCA	Trp	Ser	3	-0.31	2	-0.17
ND5	2702	GTA	ATA	Val	Met	4	0.39	1	-0.04
ND5	2703	ACC	GCC	Thr	Ala	6	0.59	2	0.24
ND5#	**2719#**	AGC	GGC	Ser	Gly	2	-0.2	2	0.13
ND5	2753	CTC	TTC	Leu	Phe	1	-0.08	4	-0.47
ND5	2754	ATT	ATA	Ile	Met	4	0.37	8	0.92
ND5	2755	AAT	AGT	Asn	Ser	1	0.1	2	0.19
ND5	2827	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND5	2847	GGC	GAT	Gly	Asp	4	0.44	3	-0.33
ND5	2847	GGC	AGC	Gly	Ser	2	0.2	2	-0.13
ND5	2851	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND5	2852	CTC	TTC	Leu	Phe	1	-0.08	4	-0.47
ND5	2898	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND5	2904	TTA	ATA	Leu	Met	3	0.24	1	-0.08
ND5	2913	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND5	2913	ACT	GCT	Thr	Ala	6	0.59	2	0.24
ND5	2916	TCT	ACT	Ser	Thr	1	0.06	1	-0.11
ND5	3012	TCC	ACC	Ser	Thr	1	0.06	1	-0.11
ND5	3061	GCA	GTA	Ala	Val	4	-0.36	1	-0.02
ND5	3080	CCC	TCC	Pro	Ser	2	0.2	2	0.22
ND5	3081	CTT	TTT	Leu	Phe	1	-0.08	4	-0.47
ND5	3090	GCA	GTA	Ala	Val	4	-0.36	1	-0.02
ND5	3093	AAT	ATT	Asn	Ile	4	0.41	6	-0.66
ND5	3125	ACA	GCA	Thr	Ala	6	0.59	2	0.24
ND5	3125	ACA	GCA	Thr	Ala	6	0.59	2	0.24
ND5	3133	GGT	GCT	Gly	Ala	8	0.85	1	0
ND5	3133	GCT	ACT	Ala	Thr	6	-0.59	2	-0.24
ND5	3136	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND5	3147	ACT	AGT	Thr	Ser	1	-0.06	1	0.11
ND5	3153	ATG	TTT	Met	Phe	3	-0.32	4	-0.39

ND5	3155	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND5	3155	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ND5	3155	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ND5	3155	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND5	3156	ATA	ACA	Met	Thr	6	-0.62	2	-0.18
ND5	3156	ACA	GCA	Thr	Ala	6	0.59	2	0.24
ND5	3156	ACA	ATA	Thr	Met	6	0.62	2	0.18
ND5	3156	ATA	ACA	Met	Thr	6	-0.62	2	-0.18
ND5	3158	TCA	CTA	Ser	Leu	4	0.44	2	0.15
ND5	3160	GCC	ACT	Ala	Thr	6	-0.59	2	-0.24
ND5	3171	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ND5#	**3173#**	GCA	ACA	Ala	Thr	6	-0.59	2	-0.24
ND5#	**3173#**	GCA	ACA	Ala	Thr	6	-0.59	2	-0.24
ND5	3179	GCC	GTT	Ala	Val	4	-0.36	1	-0.02
ND5	3179	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ND5#	**3195#**	GTA	CTA	Val	Leu	2	0.15	1	0.04
ND5	3208	ATT	GTT	Ile	Val	1	-0.02	8	0.96
ND5	3211	GCT	GTT	Ala	Val	4	-0.36	1	-0.02
ND5	3211	GCT	ACT	Ala	Thr	6	-0.59	2	-0.24
ND5	3217	GCT	TCT	Ala	Ser	6	-0.65	2	-0.13
ND5	3241	GTT	ATT	Val	Ile	1	0.02	8	-0.96
ND5	3246	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND5	3247	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
ND6	3248	ATG	GTG	Met	Val	4	-0.39	1	0.04
ND6	3254	ACT	ATT	Thr	Ile	3	0.25	6	-0.74
ND6	3257	GTA	ATT	Val	Ile	1	0.02	8	-0.96
ND6	3267	GCT	GTT	Ala	Val	4	-0.36	1	-0.02
ND6	3346	TTA	GTA	Leu	Val	2	-0.15	1	-0.04
ND6	3346	CTT	TTT	Leu	Phe	1	-0.08	4	-0.47
ND6	3354	GGC	AGC	Gly	Ser	2	0.2	2	-0.13
ND6	3360	ATA	GTA	Met	Val	4	-0.39	1	0.04
ND6	3397	GTT	ATT	Val	Ile	1	0.02	8	-0.96
Cytb	3422	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
Cytb	3436	AGT	GGT	Ser	Gly	2	-0.2	2	0.13
Cytb	3477	TCA	GCA	Ser	Ala	6	0.65	2	0.13
Cytb	3516	ATA	ACA	Met	Thr	6	-0.62	2	-0.18
Cytb	3534	AAC	AGC	Asn	Ser	1	0.1	2	0.19
Cytb	3660	ACC	ATC	Thr	Ile	3	0.25	6	-0.74
Cytb#	**3661#**	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
Cytb#	**3661#**	GCC	ACC	Ala	Thr	6	-0.59	2	-0.24
Cytb	3677	ACC	ATC	Thr	Ile	3	0.25	6	-0.74
Cytb	3682	ATT	GTT	Ile	Val	1	-0.02	8	0.96
Cytb	3722	ATA	CTA	Met	Leu	3	-0.24	1	0.08

Cytb	3723	GTT	ATT	Val	Ile	1	0.02	8	-0.96
Cytb	3761	GTA	ATA	Val	Met	4	0.39	1	-0.04
Cytb	3773	GTG	CTG	Val	Leu	2	0.15	1	0.04
Cytb	3784	GTT	ATT	Val	Ile	1	0.02	8	-0.96
Cytb	3798	GCA	GAA	Ala	Glu	1	0.09	2	-0.15
Cytb	3798	GAA	CAA	Glu	Gln	4	-0.4	1	-0.02

Table S-11. Locality information including station number, depth of haul, latitude and longitude, and time of day for samples used to prepare RNA libraries along with short-read archive (SRA) accession numbers.

Sample ID	Station	Max Haul Depth (m)	Total Length of Fish (mm)	Start Latitude	Start Longitude	Start Time (Local)	End Latitude	End Longitude	End Time (Local)	Local Date	SRA accession
1	A6-20	20	74	70.429	146.077	14:06	70.431	146.081	14:08	13 August 2013	SRX3165582, SRX3165583, SRX3165584
2	A6-1000	200	103	71.022	146.135	4:55	71.077	145.955	6:10	17 August 2013	SRX3165581, SRX3165585, SRX3165588
3	A2-1000	1000	99	70.623	142.119	8:31	70.623	142.122	8:34	18 August 2013	SRX3165577, SRX3165578, SRX3165587
4	A2-20	10	23	69.961	142.208	12:20	69.873	142.186	12:42	20 August 2013	SRX3165586
23	A1-20	20	21	69.717	141.119	17:40	69.717	141.121	17:48	20 August 2013	SRX3165580
30	MAC-1000	200	37	70.585	139.754	11:36	70.606	139.834	13:08	30 August 2013	SRX3165579

Department of the Interior (DOI)

The Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors the Nation’s trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated island communities.

Bureau of Ocean Energy Management (BOEM)

The mission of the Bureau of Ocean Energy Management is to manage development of U.S. Outer Continental Shelf energy and mineral resources in an environmentally and economically responsible way.

BOEM Environmental Studies Program

The mission of the Environmental Studies Program is to provide the information needed to predict, assess, and manage impacts from off-shore energy and marine mineral exploration, development, and production activities on human, marine, and coastal environments. The proposal, selection, research, review, collaboration, production, and dissemination of each of BOEM’s Environmental Studies follows the DOI Code of Scientific and Scholarly Conduct, in support of a culture of scientific and professional integrity, as set out in the DOI Departmental Manual (305 DM 3).